Università Ca'Foscari Venezia

Corso di Dottorato di ricerca
in Scienze e Gestione dei
Cambiamenti Climatici
ciclo 30

Tesi di Ricerca

A Statistical Framework for Modeling the Spatial Distribution and Intensity of Orographic Precipitation

Coordinatore del Dottorato
prof. Carlo Barbante

Supervisore
prof. Silvio Gualdi
dott. Stefano Materia

Dottorando
Paola Marson
Matricola 956160

Abstract

Most precipitation in the mid-latitudes is attributable to convective clouds and frontal systems. Although their development is most often not induced by the underlying terrain, the orography can play a substantial role in altering their features. A potential consequence is the modification of the spatial distribution of precipitation at the ground, which likely exhibits a highly segmented nature in both space and time.

For statistical modelling purposes, observed precipitation data are in general too coarse in space to be representative of the topographic variations of the "true" precipitation field. Aiming at describing the spatial distribution and intensity of precipitation over complex terrain, we specify a statistical model which does not solely rely on observed data but also incorporates established analytical approximations of the physical processes involved.

In particular, a 2-dimensional advection equation for the column integrated hydrometeor density is derived, by reducing the quasi-analytical up-slope model by Smith and Barstad, 2004. The advection equation contains terms that model local orographic processes causing condensation and hydrometeor production, as well as hydrometeor advection and precipitation. A Simultaneous AutoRegressive model for a latent *potential precipitation rate* has then been defined by discretizing the advection equation and perturbing it by means of stochastic noise. The moment statistics of the latent field depend on both physical quantities and unknown parameters estimated from data. Occurrences and intensities of precipitation are then derived on the basis of the latent field. An extension of the statistical model also includes a representation of large-scale precipitation.

The statistical model inherently uses atmospheric fields that can be provided by numerical climate models. Thus, the framework can be nested into climate models and provide a process-informed and spatially-consistent sub-grid refinement adjusted by observations.

The model was used to perform idealized experiments and applied to real case studies, in which showed a good agreement with observations.

Contents

Introduction

1.1 Motivation and Problem Statement

The term "*orographic precipitation*" traditionally refers to the rain and snow resulting when airflow interacts with mountains (Houze, 2012).

Most precipitation in the mid-latitudes is attributable to convective clouds and frontal systems. Although their development is not induced by the underlying terrain, the orography can play a substantial role in altering their features. The dynamic of the impinging moist flow interacts with thermodynamic and microphysical processes as well as with the large-scale flow and the height and shape of the topographic features. Moreover, as stated in Roe, 2005, orographic precipitation is a transient phenomenon, in the sense that the precipitation rates can sensibly vary as the synoptic conditions evolve during the passage of a preexisting storm.

The perturbation of the large-scale flow when it interacts with mountain ranges is visible on a wide range of length scales, from the one typical of boundary layer motions to the one of planetary waves (Smith, 1979). Consequently, the signature of the mountains on the spatial pattern of precipitation is visible from very short distances (few kilometers) to length scales greater than the one of the mountain range itself; the latter case favors or enhance the formation of some of the desertic areas on the planet (e.g Great Indian Desert, Atacama Desert in Chile, Death Valley in the USA, Piedmont in the italian Alps); at short scales, the potentail consequence is the modification of the spatial distribution of precipitation at the ground, which likely exhibits highly segmented nature in both space and time.

Observations of precipitation are usually composed of long-term records of data from sparse rain gauge stations. When from sparse recorded data one attempts to create an interpolated surface over an area of interest, the accuracy of precipitation estimates is dependent on the density and distribution of rain gauge stations. Indeed, if data are not representative of the studied process, the accuracy of the resulting interpolated fields might be questionable. The poor representativeness of observed data of the topographic variations of the "true" precipitation field may be due to different reasons. For example rain gauges over mountains are are in general too coarse in space for that scope: their spatial density can span from at least one station every $5Km$ in Switzerland to less than 1 station in $80km^2$ in known under-sampled areas as over the Himalaya. Another potential limitation is linked to how the stations are distributed, since most of the gauges are usually set at lower elevations. Because

both climate and orographic processes at high elevations are quite different than the ones at low elevations, the data are not necessarily able to capture the signal with an adequate homogeneous representation over a catchment of interest. A similar issue could arise when two mountain slopes close in space but differently exposed to the incoming moist flow have rain gauges with different density. Thus data do not necessarily contain all the signal of the precipitation process one wants to map.

For mapping orographic precipitation, both statistical approaches and analytical models were proposed in literature, the second coming from the work by Smith, 1979. Approaches from both the branches are presented and compared in section 1.2.2. These model necessarily allow a partial and approximated description of the physical orographic processes and sometimes rely on parameterization to assess and tune for introducing processes not completely understood or of a very high complexity.

The objective of this work is the definition of a statistical framework able to describe the spatial distribution and intensity of precipitation over complex terrain and to create interpolated surfaces of rainfall occurence and intensity accordingly. The model does not solely rely on observed data but also incorporates scientific knowledge on the phenomena involved. This is achieved by incorporating into the model established analytical descriptions of some of the physical processes causing orographic precipitation and, at the same time, letting unkown parameters be estimated from the observations. Combining the two approaches it will be possible to gain information from both the observed data and the knowledge on the processes. Because of the nature of processes included in the model, the framework is customized for 6-hourly to daily data.

Chapter 1 is devoted to the review of scientific background : section 1.2.1 presents the main mechanisms of orographic precipitation; section 1.2.2 reviews the existing methods for mapping precipitation over mountainous regions. Chapter 2 presents the statistical model here defined, together with some theoretical discussion and extentions of the model. Chapter 3 collects experiments carried out with the statistical model: section 3.1 presents experiments on a set of idealized terrains, while section 3.2 shows case study on real data over California. A discussion and some conclusions are reported in chapter 4.

1.2 Scientific Background

1.2.1 Mechanisms of Orographic Precipitation

The categories of storms responsible of most of the precipitation in the atmosphere are convective clouds and frontal systems in the mid-latitudes and tropical cyclones in the tropics. These formations originate due to atmospheric synoptic mechanisms not related to the features of the underlying terrain, with the exception of rare and occasional episodes. Convective storms are formed drawing energy from the vertical stratification of temperature and water vapor in the atmosphere. Deep and broad convective cores might be enhanced or modified near and over mountains. That is especially true in warm seasons by the control exerted by the diurnal cycle. Frontal systems draw their energy from the horizontal variation of atmospheric temperature, when warm and moist air from lower latitudes enters extra-tropical cyclones and rises as *conveyor belts* over cold and warm fronts in the mid-latitudes producing clouds and precipitation. Orographic enhancement of cyclonic rainfall can occur on the windward sides of the mountain ranges when the storm is already mature and the modification of the fallout depends also on the maturity stage of the storm itself when it passes the topographic formation. Tropical cyclones draw their energy from the latent heat present in the upper levels of the oceans. Occasionally precipitation may also be caused solely by orographic mechanisms when the wind blows over rugged terrain apart from the occurrence of one of the synoptic storm categories. It happens for example when moist trade winds blow over the mountainous islands of Hawaii.

In general, the mountains interact with the atmosphere affecting flow convergence, flow dynamics, cloud microphysics, the humidity and thermodynamics, depending on the shape and relative position of the topographic barrier and the nature of the incoming formed weather system and low level flow.

The influence of topography on precipitation can be attributed primarily to localized disturbances of the vertical structure of the atmosphere. Such disturbances result from orographic formations acting either as

- an obstacle for the impinging moist air triggering dymanic and thermdynamic processes in the flow.

- elevated heat sources and sinks which in general produce diurnal cycles in the vertical motions.

- concentrated source of roughness in rare cases like remote isolated islands or isolated peaks.

Thermodynamics

The vapor contained in an air parcel reaches saturation if its partial pressure reaches the saturation vapor pressure. Because the last decreases exponentially with temperature and hence with height, condensation might be triggered by vertical ascent of moist air when it is forced to pass over a topographic obstacle. Figure 1.1 from Sumner, 1988 describes the most important features of the thermodynamics involved: the first panel shows the orographic uplift and cloud formation; the second panel is a pseudo-adiabatic chart.
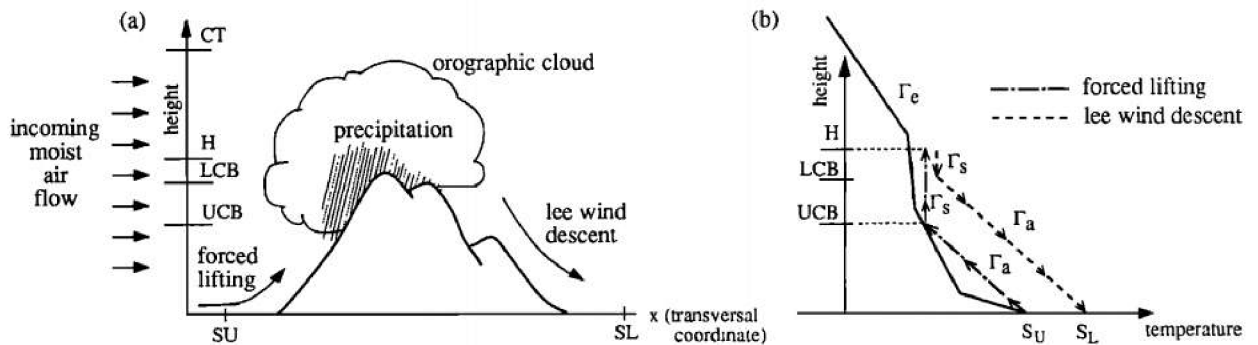

Fig. 1.1: Figure from Sumner, 1988. Left panel: representation of the thermodynamic processes leading to orographic precipitation; rigth panel: pseudo-adiabatic chart relative to the situation depicted in thepanel on the left.

An unsaturated air parcel contained in the flow passing over a topographic obstacle ascends adiabatically on the upwind slope of a mountain barrier (SU in the figure). As the air parcel rises, it cools at a rate determined by the lapse rate Γ_a. Condensation is triggered when the air parcel can no longer hold water in the vapor phase because the parcel's water vapor pressure which is decreasing while the parcel rises reaches its saturation value or, in other words, the rising parcel reaches its lifting condensation level (LCL in the figure) and the cloud base level (UCB) is formed. Above this altitude the air parcel is saturated.

When condensation is activated, a new cloud is created, or a preexisting one is enhanced. From this point the thermodynamic changes take place at the saturated adiabatic lapse rate

$$\Gamma_s = \Gamma_a(1 + \epsilon) \tag{1.1}$$

where ϵ is the ratio of the heat released because of condensation to the corresponding change in internal energy, in a hydrostatic atmosphere and under pseudo-adiabatic conditions.

Part of the condensed water can be lost by precipitation, that mainly occurs in the region which spans from the cloud base (UCB) to the mountain top (H). As the

air parcel passes the mountain peak and moves downward along the lee slopes, it remains saturated until it reaches the lee side cloud base level (LCB); this level is at lower elevation than on the windward side where the parcel still retains all ts original moisture content. Below the LBC, the parcel becomes unsaturated again, and its descent proceeds adiabatically.

The differences in the process in the upwind and downwind side of the orographic formation induces a precipitation divide in the two regions which is known as *Rain Shadow Effect*. It is responsible of the formation of the major desertic areas on the continents as well as contributes greatly to the spatially-segmented nature of precipitation at a more local level in the mountain valleys of the same range.

Flow dynamics and static stability

The impinging airflow responds differently mostly depending on three elements:

(a) Upstream flow velocity   (b) Static stability   (c) Mountain shape

The three information are often combined in the so called *Non-dimensional Mountain height - Inverse Froude Number*:

$$M = \frac{N_m h}{U} \tag{1.2}$$

where U is upwind cross-barrier flow, $N_m = \frac{gT}{\Gamma_d - \gamma}$ is the moist Brünt-Vaisala frequency, h is the maximum height of the terrain and Γ_d and γ are the dry and environment lapse rate respectively.
When the ratio $\frac{U}{N_m h}$ is large, the airflow easily rises over the terrain, and when it is small, the oncoming airflow may be blocked, and its horizontal trajectory may be deflected. Moreover, the ratio is a measure of the importance of non-linear regimes in the flow, so taht when the ratio is greater then one, the linear theory for density-stratified fluid might not be appropriate. (Smolarkiewicz and Rotunno, 1989)

Stable flow: Stable moist air impinging a topographic barrier generally follows the terrain upward. The vertical motion produces a cloud in the windward side of the mountain or strengthens a preexisting larger cloud entity (e.g. one associated to a frontal system), while on the lee side the cloud is evaporated. This process applies to terrain of any size.

The stable cross-barrier flow, furthermore, sets off *gravity waves* motions. The result is the presence of two regions of vertical motion: the upslope flow in the windward slope and a vertically propagating gravity waves on the lee side and over the crest,

where a cloud might be produces aloft. Moreover, vertically propagating wave motion interacts and/or favors precipitating convective clouds on the lee side. Fig. 1.2 from Jiang and Smith (2003) shows the vertical cross section of the steady-state solutions simulated by the ARPS numerical model with control parameters Brunt-Väisälä Frequency $N_d = 0.011s^{-1}$, horizontal wind $U_0 = 10m/s$. The vertical velocity is contoured, and regions of upward and downward motion appear induced by gravity waves.
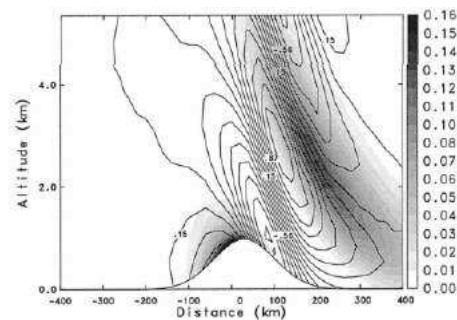

Fig. 1.2: Figure from Jiang and Smith, 2003 . Vertical cross section of the steady-state solutions simulated by the ARPS numerical mode for a statically stable flow. The vertical velocity is contoured, the mixing ratio of snow is shaded.

Unstable flow : Air ascending the terrain can overturn on a sub-barrier scale as it rises. In case of intense unstable conditions, the resulting convective clouds could reach great heights ($\sim 10Km$) and be triggered within pre-existing cloud systems. Moreover, if the side of the mountain range is jagged, intense convective clouds can form and water precipitate over the upslope side of each consecutive peak.

Blocking : In case of high values of the Inverse Froude Number, when the air upstream of a mountain barrier is strongly stable or has a weak cross-barrier flow component, the air is blocked and does not rise over the terrain. One result is that the effective obstacle to pass is moved upstream and the air ascends well ahead the mountain barrier, rising over the low-level blocked flow.

Sometimes the blocking is only partial and can produce an acceleration in the downward flow in the lee side of the mountain. The flow returns to an equilibrium in the form of hydraulic jump which creates region of air ascent and precipitating cloud downstream of the barrier.

Geometry of the mountain ranges and microphysics

Microphysical processes govern the transition from the condensation of water vapor to water drops and ice crystals with precipitable size. These processes require time, so that time scales of particles growth are linked to the time scales of flow passing a

mountain range. Moreover, the height, steepness and width of the mountain range affect and limit the locations where the falling particles can land.

For a more comprehensive and deep review of orographic precipitation mechanisms, the reader is referred to Houze, 2012, Colle et al., 2013, Colle, 2004, Roe, 2005 and references therein.

1.2.2 Models for the Spatial Distribution of precipitation over complex terrain

Two main approaches are usually used to estimate the spatial distribution of orographic precipitation: the interpolation of rain gauge station data and the so called up-slope methods. Both methods have advantages and disadvantages, as well as the ability of capturing different features of the process they aim at modeling and estimating.

Methods belonging to both groups will be presented and critically compared in the following paragraphs.

Statistical approaches

The available observational data of precipitation are usually long-term records from sparse rain gauge data. In order to extrapolate information on un-sampled locations or grids points , deterministic methods (e.g. Thiessen Polygon) as well as geostatistical interpolation methods (e.g. Kriging techniques) have often been used. More recently, multiple regression approaches were introduced allowing the inclusion of secondary aspects (altitude, exposition to the wind) to improve the estimates.

A general form of *interpolation function* is common to most of the interpolation methods: let us consider the point s_j in the 2D-space, its pair (x_j, y_j) of coordinates and a function $h_j = h_j(x_j, y_j)$ which denotes the observed process of interest at the sampled $j = 1, ..., n$ points. An estimate of the process h_0 interpolated at any un-sampled site $s_0(x_0, y_0)$ can be represented as a weighted linear combination of the observed values h_j on the other n sites:

$$h_0 = \sum_{j=1}^{n} w_j h_j$$

where w_j is the weight attributed to the sampled point j.

Most of the interpolation techniques can be seen as a particular case of the previous model and differ on how the weights w are found and attributed to the points.

Deterministic methods such as Thiessen Polygon and Inverse Distance Weighting are classical methods commonly used in hydrological models. They don't include any specific attention to the topographic influence and they constitute an established benchmark in comparative studies of statistical interpolation techniques.

The *Thiessen Polygon* is a deterministic method for interpolation first introduced by Thiessen, 1911. The author recognizes that precipitation varies considerably also

over quite limited distances. When one attempts to reconstruct average values of rainfall over an area where observations are available only on sparse and unevenly distributed gauges, the amount of rain recorded at any station should be representative only for the region enclosed by an imaginary line drown midway between the station under consideration and the surrounding stations. The consequence is that the area is divided into polygons of influence and each interpolated point takes the value of the closest sampled point. Besides the advantage of the simplicity of the method, the estimation is based only on one measurement and discontinuities are created on the edges of the polygons. Recalling the general model form, the weight w_j is equal to one if the point j is the closest to the unsampled point on which the value h is estimated or zero otherwise.

The *Inverse Distance Weighting (IDW)* algorithm is based on the idea that the value at a point is more influenced by closer points than by points farther away. The method permits to calculate the value at un-sampled point by the weighted average of data observed at surrounding points. The weights are determined according to the inverse of the distance between the points. For determining the value h_0 on point s_0 the weight w_i attributed to the generic sampled value at point s_i can be determined as:

$$w_i = \frac{\frac{1}{|D_0i|^d}}{\sum_{j=1}^{ns} \frac{1}{|D_0j|^d}}, \quad d > 0 \tag{1.3}$$

where D_0i is the distance between the sample and un-sample point s_0 and s_i and d controls for the distance - decay effect. Dirks et al., 1998 found that for precipitation, depending on the temporal scale considered, the distance of influence varies. In particular they use a value d equal to 1 for yearly time step, 2 for daily and monthly time steps and 3 for hourly data minimizes the interpolation errors.
The method was extended by Lu and Wong, 2008 who proposed an adaptive version, where the inverse-distance weight is modified by means of a decay parameter that is allowed to vary over the study area. Although IDW is a relatively simple deterministic interpolation method, which provides adaptable weights for sensible local interpolations, the choice of the weighting function is arbitrary.

Geostatistical Interpolation Methods use statistical properties inferred from data and geographical information to interpolate recorded data.

`Kriging methods` were at first developed by Matheron, 1971 who introduced the concept of *regionalized variables* to estimate areal averages considered as realizations of stochastic processes. These methods were subsequently applied to the interpolation of precipitation stations (e.g Tabios and Salas, 1985). Like the two methods previously described, the Ordinary Kriging relies on the concept the spatial points influence each other according to their proximity and provides an estimate on an un-sample location as linear relationship of the value at the surrounding points. The

weights w are determined taking advantage of the mean and spatial correlation structure found in the data. This is achieved examining the strength of the covariation of the process between two sites as function of the pair-wise distance between the same two sites, by means of *semivariograms* estimated from the data.

In particular, defining the *semivariogram* for the points s_i, s_j as $\gamma(D_{ij}) = \frac{1}{2}var[h_i - h_j] = \sigma^2 - cov(D_{ij})$ where D_{ij} is the distance between s_i and s_j and σ^2 is the marginal variance, the weights w for interpolating h on the site s_0 can be determined minimizing the variance of the errors:

$$Var[\epsilon_0] = \sigma^2 - 2\sum_j w_j[\sigma^2 - \gamma(D_{0j})] + \sum_i \sum_j w_i w_j[\sigma^2 - \gamma(D_{ij})]$$

Kriging methods require second order stationarity (homogeneity of the mean and of the covariance as function of distance). The reader is referred to classical textbook such as Cressie, 1993 for a more in-depth description of this methodology.

The use of ordinary Kriging over mountainous regions is questionable. Indeed, Kriging uses only the information carried by data: in its ordinary formulation, it does not account for any external aspects (such as elevation) except indirectly for the part of their influence reflected and represented in the data. However, this means that the estimate on an unsampled point is likely to be mistaken if the surrounding points fall at different elevations (Tabios and Salas, 1985). On this regard, Dingman et al., 1988 performed a linear regression on precipitation vs. elevation, then they performed kriging on residuals after subtracting the regressed elevation effect. Ly et al., 2011b studied the effect of using different parametric models for the semivariogram when performing kriging for the interpolation on 1 Km^2 regular grids of daily precipitation data.

Extensions to Ordinary Kriging have been introduced and their use proposed for the interpolation of precipitation data over complex terrain with the primary aim of improving the estimates including the effect (or covariation) of auxiliary variables (e.g. elevation).
In Phillips et al., 1992 the authors compare two extensions of kriging such as ordinary kriging on elevation-detrended data, proposing to detrend the data using spherical, exponential and gaussian functions of altitude. They also proposed the interpolation method Cokriging (Goovaerts, 1997) using elevation as auxiliary variable.

Cokriging is a multivariate extension of kriging that calculates estimates of a poorly sampled variable with help of a better-sampled co-variable. The determination of the weights w is then based on a *cross-variogram* model for the combination of both variables.

Luo et al., 2011 showed Cokriging may be advantageous when the auxiliary variable is highly correlated ($r^2 > 0.5$) with the primary variable and the former is over-sampled, while could lead to poorer results when the correlation with elevation is weak.

On the correlation between precipitation data and elevation, Daly et al., 1994 reviewed previous works which assessed with data the general increase of climatological precipitation with elevation in different areas. The author also reviewed the assessment of the relative spatial location of the most intense rainfall rate with respect to the shape of the mountain. Although it is not possible to generalize the location where the maxima occur for mapping purposes, in some of the reviewed works precipitation maxima in the midlatitudes are found at or near the crests, but maxima can also occur below the crests on the windward slope in subtropical regions (e.g Hawaii islands) and in the mid-latitudes in high or broad mountain ranges. In these conditions the clouds discharge precipitation before reaching the peak (e.g Sierra Nevada). Moreover, in narrow but very steep mountains, the condensed water can be advected to the lee side before precipitating and create a climatological maximum of rainfall downwind the crest. Similarly, Frei and Schär, 1998, studying annual means on the eastern european Alps, found precipitation is more intense at lower levels in high mountains because of the exponential relationship between saturation vapor pressure and temperature (and hence elevation). Given this complexity, Houze, 2012 noted that interpolation schemes that automatically set precipitation amount proportional to terrain high might be in error especially in areas of high terrain.

In this thesis a very similar result is found contrasting elevation and precipitation over the Oregon and Washingthon states for the climatological 1985-2015 daily precipitation, as shown in Fig. 1.3 The analysis was repeated for both occurrence (estimated probability) and intensity dividing the entire region into three subregions with slightly different mountain shapes . The right panels in the figure show the sub-domain on which the analysis was performed (coloured area). In each sub-domain the longitudinal profile of the mountain range was built considering the latitudinal mean altitude. The left panels present the terrain profile (black) and the probability of rainfall (orange) for the three sub-domain. Similarly, the panels in the middle column show the terrain profile and the daily mean amount of precipitation (orange). Almost identical patterns are also found considering seasonal and annual values.

Groisman and Legates, 1994 also found that unreliable data in complex terrain at high altitudes could also lead to estimate biases as large as 25% where snow accumulates. On the same aspect, Daly et al., 1994 reviewed functional forms used to relate altitude to precipitation intensity, such as linear, log-linear and exponential.
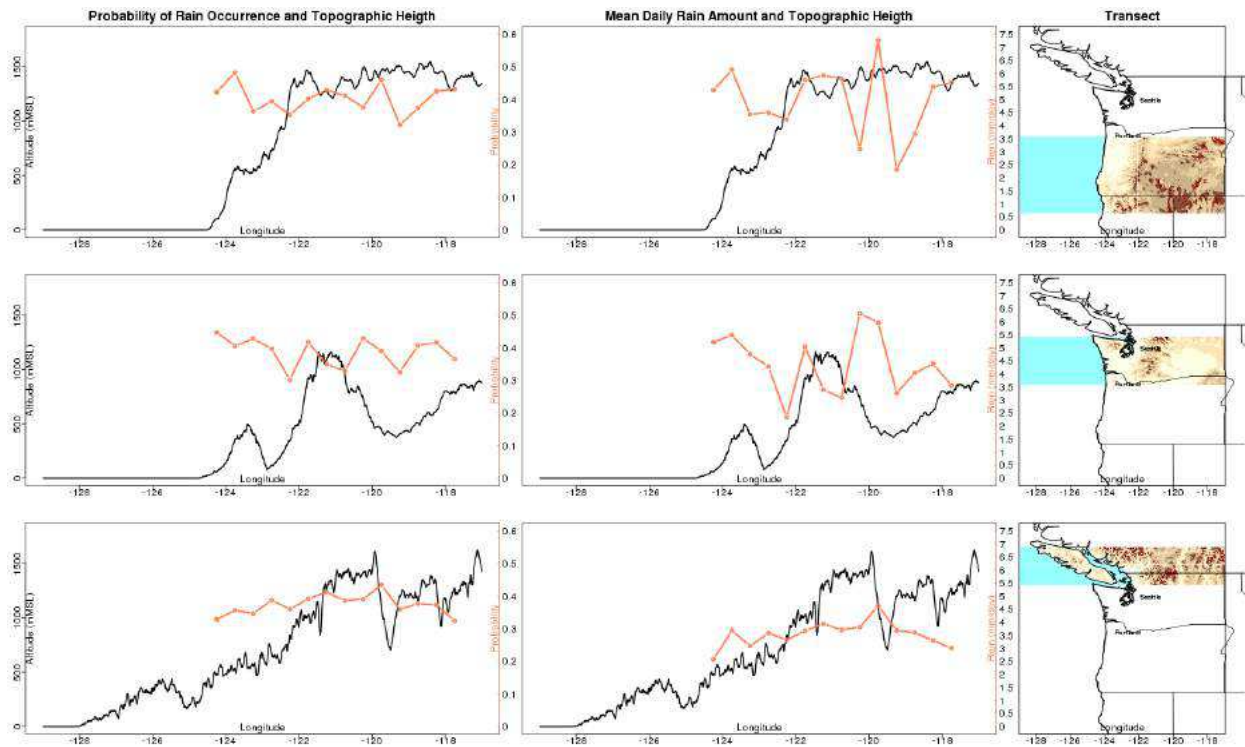
Fig. 1.3: E-W sections of Cascades (OR and WA, USA) and observations of probability of precipitation occurrence and precipitation intensity.

Among statistical methods that incorporate secondary information, Goovaerts, 2000 instead compares three multivariate geostatistical algorithms for incorporating elevation into the spatial prediction of precipitation, with an application in Portugal. In particular the author used 1) a simple kriging letting the mean vary locally as a linear relationship of elevation; 2) kriging with external drift: after deriving the locally varying mean as in 1), a simple kriging is performed on the residuals; 3) Co-located kriging (Goovaerts, 1998): as opposite to the first two methods, the latter incorporates altitude also in the surrounding area of the point of interest.

Modification of the standard procedures just presented were introduced by Ly et al., 2011a who compared seven semi-variogram models (logarithmic, power, exponential, Gaussian, rational quadratic, spherical and penta-spherical) fitting daily data with a particular attention in avoiding predicting negative precipitation, and they found the Gaussian model was the best fit most frequently.

Schiemann et al., 2011 combined gage data with radar data and proposed the use of a non-parametric *semi-variogram* that was then tested using the different kriging approaches previously mentioned.

Many uncertain factors enter the choice of the interpolation method to use. Regarding the uncertainty inherent in the time dimension, two works are reported here in which the uncertainties in time scales and time spans are analyzed:

Chen et al., 2017 pointed the attention on the peculiarity of the time scales of the interpolated data: they recognized that the use of different interpolation schemes in the same catchment may cause large differences in spatial distribution of rainfall, and that methods can perform differently on the same spatial domain depending on the temporal scale considered. For this reason they proposed an improved regression-based scheme using principal component regression with residual correction: they added topographic factors as covariates based on station latitude and longitude, elevation, slope, and slope aspect. They then applied an inverse distance weighted regressions on the residuals. Moreover the method uses slightly different data processing methods for each of three time scales, namely hourly, daily and annual. The author compared the proposed method to an inverse distance scheme and multiple regression for the three time scales on the mountainous catchment of the Fuhe River in southeastern China.

Focusing more on the time span of the analysis, Lloyd, 2005 compared (i) moving window regression, (ii) inverse distance weighting; (iii) ordinary kriging, (iv) simple kriging with a locally varying mean and (v) kriging with an external drift for monthly precipitation data in Great Britain. Methods (i),(iv),(v) make use of elevation as secondary variable. They applied all the methods to predict values for each month in 1999 and concluded that in different months different methods of interpolation perform best.

Some studies focused on interpolating simultaneously both temperature and precipitation (e.g Buytaert et al., 2006, Daly, 2006). Indeed, single regression functions may not accurately represent spatially varying meteorological variables across large regions (Daly, 2006). In this regard, Tobin et al., 2011 proposed a variant of Kriging with external drift using an unbiased robust anisotropic *variogram* and using variables from a 6-h numerical weather forecasts model and elevation data as covariates. They estimated simultaneously both precipitation and temperature. This approach allowed to provide instantaneous lapse rates which better capture snow/rainfall partitioning.

Another class of approaches are those based on **Polynomial Regression - Multiple regression** . The first is a type of regression that prescribes an *r*-th order polynomial relation between one or more predictors and the dependent variable; the second is a regression model with more than one predictor or independent variable. The peculiarity of this class of approaches is their ability to use information from many secondary variables at the same time, regardless their geostatistical attributes. In the literature,many authors introduced sophisticated variables as covariates attempting to represent the effect on precipitation of geographic, topographic and large-scale climatic features at the same time.

Goodale et al., 1998 used a second-order polynomial least-squares regression for monthly precipitation, monthly averaged maximum and minimum daily tempera-

ture, and monthly averaged daily solar radiation, using position and elevation as independent variables.

Ninyerola et al., 2007 developed a multiple regression, combined with a residual correction method to obtain a predicted surface on 200m resolution over the Iberian peninsula. The method is a combination of statistical (multiple regression) and spatial interpolation (splines and inverse distance weighting) tools, the second approach used as a further mapping on the residuals of the first model. The covariates used are altitude, latitude, continentality, terrain curvature and solar radiation. The authors also presented an interesting discussion on the choice of non-frequently used geographical and climatological variables as covariate of the model.

Agnew and Palutikof, 2000 proposed a regression-based approach to build high-resolution (1km) maps of mean seasonal temperature and precipitation for the Mediterranean Basin. The covariates used for model development include: longitude, latitude, elevation, distance from the nearest coast, direction to the nearest coast, slope, aspect, and the ratio of land to sea within given radii. Feidas et al., 2014 suggested a methodology for mapping seasonal and annual air temperature and precipitation in Greece using several topographical and geographical variables (location, elevation, distances from the coasts, slope, Sea-to-land ratio, Normalized difference vegetation index (NDVI)). The model shows different levels of accuracy in each season.

As pointed by Chen et al., 2017, studies using these approaches focused only on monthly or seasonal data, making the evaluation of their performance not straightforward for other time scales. Moreover, when many independent variables are used, multicollinearity may exist among them, even if some studies used stepwise regressions to select the most informative variables to use.

Among approaches introduced in the contest of algorithmic modeling, **Support Vector Machines (SVM)** (Vapnik and Vapnik, 1998) and **Neural Networks** were also applied for mapping precipitation respectively by Li et al., 2011, who also introduced a combined use of SVM and Kriging and by Di Piazza et al., 2011 for mean annual and monthly rainfall data in Sicily (Italy).

Other models proposed in literature do not fit inside any broader class of statistical approaches. It is meaningful to mention here other interpolation methods that account in various way for the effect of elevation: Şen and Habib, 2000 created four different precipitation categories according to relative positions and elevation and by means of a cumulative semi-variogram; Gottardi et al., 2012 used a linear relationship of elevation which strength on each day depends on the concurrent weather pattern; other models make predictions depending on the dominant wind direction and the spatial scales at which topography influences precipitation (Meersmans et al., 2016) or allow for an auto-search of most-informative climatic and topographic characteristics (Guan et al., 2005).

A spatio-temporal **Autoregressive Model** for daily precipitation was introduced by Sigrist et al., 2011. For capturing the spatial dependence the authors used an approach based on an autoregressive convolution and they also modeled temporal dependencies. A rich literature exist for spatio-temporal precipitation data, but its description is beyond the scope of this review.

Hutchinson, 1998 introduced a **Thin Plate Smoothing Spline** for the spatial interpolation of annual mean rainfall. The author also incorporated a continuous, spatially varying, dependence on appropriately scaled elevation, that made a dominant contribution to the surface accuracy.

PRISM (Precipitation-elevation Regressions on Independent Slopes Model, Daly et al., 1994) assesses the elevation-precipitation relationship and the spatial scale at which the topography influences the spatial distribution of precipitation. The characteristic feature of this model is the division of the spatial surface into *facets* with same or very similar topographic characteristics and the definition of *topographic regimes*, in term of orientation and steepness of the slopes.

Two common limitations are evident in many of the methods mentioned in this section: they do not include any natural or physical scale, so that, as an example, wide mountains and narrow mountains are treated identically. Moreover they rely on the indirect contribution of topographic and climatic aspects, not considering explicit descriptions of known physical mechanisms.

Up-slope models

The *Up-slope models* constitute a category of analytical or quasi-analytical models in which the terrain slope and wind speed are used to estimate the condensation rate above the terrain due to forced uplift of moist flow impinging a topographic barrier.

Early and grounding contributions in this direction are the models by Collier (1975), Rhea (1977), and Smith (1979). An assumption common to these models is that the water that condenses for forced uplift falls immediately to the ground. The consequence is that physical scales in both the space and time domains are not directly taken into account.

Several quasi-analytical models have instead included scale-dependent processes:Hobbs et al. (1973) and Bader and Roach (1977) have described raindrop/snow fall speeds and washout of cloud water.

Alpert and Shafir (1989), Sinclair, 1994, and Smith, 2003 have included the advection of condesed water of hydrometeors in their upslope models. These descriptions allow the condensed water to drift downwind improving the agreement with observed rainfall patterns.

Upslope-advection models such as Smith, 1979 assume that the terrain-induced vertical air velocity penetrates upward through the moist layer without changing velocity. It is known from mountain wave theory that such motions may either decay or oscillate with altitude, depending on the horizontal scale and aspect ratio of the terrain, static stability and wind speed. In either case, the upslope estimates may exceed the actual condensation rate.

A second possible issue is related to the leeside descent: in upslope models, it is usually assumed that only upslope regions influence precipitation. With instantaneous conversion and fallout, the windward slope receives rain while the flat and downslope regions are dry. When time delays and advection are included like in Smith, 2003, the water condensed for uplift motions is distributed downstream, with no regard for local terrain. This approach neglects the evaporation of cloud water and hydrometeors caused by descending air and the total precipitation is overestimated by this assumption. Smith and Barstad, 2004 adjusted the approach for evaporation in the lee side using negative terms of evaporation in the leeside, but still does not account for the possible evaporation of the advected condensed water. Alpert and Shafir (1989) introduced in this regard an adjustable coefficient to reduce the prediction into a reasonable range.

The errors induced by these two issues are strongly scale dependent. For smooth hills with scales of approximately 100 km, the upslope estimates are usually quite reasonable. When the terrain is rising and falling with scales of 20 km or less, both assumptions fail and models can overestimate the total precipitation by a factor of 5 or larger.

Methods

In this chapter the statistical model derived for mapping precipitation over mountainous terrain is presented. It was defined after deriving a reduced version of the up-slope time-delay model by Smith, 2003 introduced in the section 1.2.3. Hereafter the up-slope model is more extensively presented as most of its features and skills are still found in the statistical model. Then the derivation of the latter as well as the description of its properties and estimation procedures for its unknown parameters is given. The chapter is completed by sections suggesting additional extensions of the model and special cases of application.

2.1 Smith's up-slope time-delay model

Smith, 2003 and Smith and Barstad, 2004 introduced an advection equation model for the prediction of precipitation over mountainous terrain.

The authors begin postulating a pair of steady-state advection equations for the following two quantities on the generic point (x, y) in the 2D-space:

- $q_c(x, y) \left[\frac{Kg}{m^2} \right]$ vertically integrated cloud water density

- $q_s(x, y) \left[\frac{Kg}{m^2} \right]$ vertically integrated hydrometeor density .

$$\frac{Dq_c}{Dt} = \mathbf{U} \cdot \nabla q_c = S(x, y) - \frac{q_c}{\tau_c} \tag{2.1}$$

$$\frac{Dq_s}{D_t} = \mathbf{U} \cdot \nabla q_s = \frac{q_c}{\tau_c} - \frac{q_s}{\tau_f}$$

(2.2)

Reading all the elements included, the core of the model is the representation of the up-slope mechanisms in the term $S(x, y) \left[\frac{Kg}{s \cdot m^2} \right]$, which represents a distributed source of condensed water arising from the forced ascent of moist air over the terrain. In particular, in this term, the horizontal wind, measures of moisture content and static stability as well as the terrain geometry are used to estimate the rate of cloud

production. Many shapes of S are possible, and the authors introduced some of them in Smith and Barstad, 2004.

One possible source of condensation is the following:

$$S(x, y) = \frac{\Gamma_m}{\gamma} \rho \cdot q_{VSAT0} \cdot \vec{U} \cdot \nabla h(x, y) \qquad (2.3)$$

where:

- $\vec{U} \cdot \nabla h(x, y)$ is the vertical velocity induced by topography (+ / -);

- $q_{VSAT0}(x, y)$ is the Saturation specific humidity at the ground

- Γ_m is the moisture lapse rate;

- γ is the environment lapse rate;

S takes positive values in region of air ascent and negative values in regions of descent, where its values are interpreted as evaporation rates. The measure of static stability in the ration $\frac{\Gamma_m}{\gamma}$ does not have any direct effect on the horizontal flow, meaning it does not deflect it in case of high stability. At the same time it enhances the condensation rates when it is greater than one and thus widens the area where significant condensation occurs.

The key and distinguishing element of this up-slope model is however the inclusion of the *characteristic time scales* τ_c, τ_f: the author states that from the condensation of water vapor to the fallout of rain, a certain amount of time is required for microphysical processes to take place. The two time-delays then quantify the time required by cloud microphysical processes for the conversion of cloud water to hydrometeors (τ_c) and for the hydrometeors fallout (τ_f). The inclusion of the time delays has the immediate consequence of allowing for the advection of the condensed water and the formed hydrometeors before reachinf the ground, which implies precipitation is distributed downwind of the cloud sources over a distance given by the time delays themselves and the wind speed. Moreover, the time parameters act as indirect control of the *precipitation efficiency*. The two time scales are not known, but analysis by Jiang and Smith (2003) suggested a feasible range for both τ_f, τ_c as $[200s, 2000s]$, and the two are usually set equal to each other ($\tau_c = \tau_f$) in applications.

The last term of the first equation $\frac{q_c}{\tau_c} \left[\frac{Kg}{s \cdot m^2} \right]$ represents the rate of conversion from cloud water to hydrometeors, while the last term of the second equation $\frac{q_s}{\tau_f} \left[\frac{Kg}{s \cdot m^2} \right]$ is the precipitation rate.

The up-slope model benefits from properties that make it suitable for real applications and that are still found in the statistical model.

- It is Linear and analytically tractable ;

- It is applicable to actual complex terrain: without the requirement of representing the terrain by idealized shapes. Moreover the shape of the real terrain becomes one of the natural length scales that can be included in the model;

- It is applicable to arbitrary wind directions: the model can be directly used in real applications without the need of rotating the coordinate axes;

- It reduces to the classical upslope model at the limit of short time delays τ;

- It includes all the basic physical elements of orographic precipitation mechanisms: airflow dynamics, condensed water conversion, advection and fallout, and downslope evaporation, leading to a theory of precipitation efficiency.

The model also has some limitations:

- It only partially accounts for flow blocking or very unstable conditions: the static stability influences the intensity of the vertical motion but does not modify the horizontal trajectory of the flow, which may happen in case of blocking.

- It is vertically integrated: it cannot account for vertical variations of the atmospheric variables included.

- It uses constant time-delays over the area, thus being unable to describe more complex variations in the microphysics.

- Because it uses steady state equations there isn't a real temporal resolution in the model.

- There is no distinction among ice and water, altough Barstad and Schüller, 2011 proposed an extension accounting for two vertical layers at this scope.

The statistical model defined in this thesis is built from a reduction of the system 2.1 by Smith (2003). The model will be described in the following section 2.2, but it is worth it to mention here that it is built on the analytical core of Smith (2003) model, yet it has more flexibility in dealing with some the mentioned limitations.

2.2 Derivation of the Statistical Model

The system of two equations specified in the model by Smith, 2003, as presented in the previous section, allows the representation and parameterization of both the processes that lead from condensation to the formation of hydrometeors and from the formation of hydrometeors to drops fallout.

The system was here simplified postulating a *single advection equation for the column integrated cloud water density* q. This simplification was possible and results in a still valid description because in the original Smith's model, the two processes don't interact but in a temporally consecutive way and moreover because using a statistical approach, precipitation is a quantity more commonly observed.

The proposed governing equation becomes:

$$\frac{Dq}{Dt} = \mathbf{U} \cdot \nabla q = S - \frac{q}{\tau} \qquad (2.4)$$

and specifically for the case of a 2-dimensional spatial domain reads:

$$\frac{Dq}{Dt} = u\frac{\partial}{\partial x}q(x,y) + v\frac{\partial}{\partial y}q(x,y) = S(x,y) - \frac{q(x,y)}{\tau} \qquad (2.5)$$

where now, following Jiang and Smith (2003) but considering here a single equation, the parameter $\tau \subset [400s, 4000s]$ parameterizes the time required for the condensed particles to precipitate, $U[m/s]$ is the horizontal wind that advects the cloud water during the time given by τ, $S(x,y)\left[\frac{Kg}{s\cdot m^2}\right]$ is a distributed source of condensation arising from forced ascent over the mountain slopes of the flow and the last term on the right hand side $\frac{q(x,y)}{\tau} = p^*(x,y)\left[\frac{Kg}{s\cdot m^2}\right]$ is again the precipitation rate.

Equation 2.5 can be conveniently written for the precipitation p^* term $\frac{q}{\tau} = p^*$:

$$\frac{Dp^*}{Dt} = u\tau\frac{\partial}{\partial x}p^*(x,y) + v\tau\frac{\partial}{\partial y}p^*(x,y) = S(x,y) - p^*(x,y) \qquad (2.6)$$

Because the source function of cloud water S as presented in 2.3 takes values on R, also p^* does. Since we will never observe a negative value for rainfall rate, p^* would rather be interpreted as a *latent* tendency rate of precipitation, whereas the values of rainfall rate / occurrence actually observed as its indicators.

2.2.1 Case 1: Precipitation Data on a Regular Grid

Equation 2.6 was then discretized using a *forward differences scheme*. This method is often referred to as *Upwind differencing*, and this choice appears natural noticing

the signal of p^* can only be propagated along the direction of the wind. In other words, the value of a field p^* at a point x_i is not affected by the flow nor by the value of p^* downwind x_i, and the signal present in x_i should instead flow farther away following the wind. The upwind scheme, moreover, produces values that are not unstable, meaning that the scheme is monotonicity preserving and does not produce local minima or maxima or fast oscillating values in the field as it could occur using *centered differences schemes*.Because the advection equation used for deriving the statistical model is a steady state advection equation, one could reasonably argue that the problem moves in a pure spatial setting and that schemes based on centered differences would be at least equally preferable. Instead, the issue on the discretization of advection equation still holds. Indeed, even in a pure spatial setting, what drives the instabilities is not considering the natural propagation of the signal downwind. Accounting for opposite contributions from both downwind and upwind sites, as it is the case in centered differences, would lead to unphysical solutions and conflicts and to the possible failure of the numerical scheme. Moreover, the truncation local error, namely the amount of error produced in one step of a numerical approximation because of the truncation, is smaller in unilateral schemes.

For positive components u, v of the horizontal wind U the following discretization is obtained, defining a *unilateral process*:

$$u\tau \frac{p(x,y) - p(x - \Delta x, y)}{\Delta x} + v\tau \frac{p(x,y) - p(x, y - \Delta y)}{\Delta y} = S(x,y) - p(x,y) \quad (2.7)$$

It is possible to perturb the previous differences by means of a stochastic noise η, making the solution of the difference equation a stochastic spatial process $\{p^*(x,y) : (x,y) \in D\}$ as well defined over a *finite regular lattice D*.

$$u\tau \frac{p(x,y) - p(x - \Delta x, y)}{\Delta x} + v\tau \frac{p(x,y) - p(x, y - \Delta y)}{\Delta y} = S(x,y) - p(x,y) + \eta(x,y)$$
$$(2.8)$$

Assuming $\{p^*(x,y) : (x,y) \in D\} = \{p^*(s_i) : i = 1, ...n\}$ is defined over a finite subset composed of n nodes of the lattice D on the plane, we can let η be a random vector following a *n-dimensional* zero mean and zero pairwise covariance gaussian distribution $\eta \sim N(0, \Lambda)$ where Λ is a diagonal variance matrix and let $\eta(x,y) \sim N(0, \sigma^2(x,y))$ be the univariate marginal distribution for the generic point (x,y) on the lattice. As generally assumed in literature, also in this thesis the case of homogeneous variance $\eta \sim N(0, \Lambda), \Lambda = \sigma^2 I$ is the only one considered, but other choices are allowed if knowledge about how to let the stochastic noise vary in space were available.

The noise added can reflect both the scientific uncertainty of the deterministic relation used and random impulses in the system with a lower order of magnitude with respect to the main signal.

Writing the previous expression in matrix notation makes the understanding of the *moments* of the defined stochastic process straightforward:

$$p^* = Ap^* + Bp^* + Cp^* + S + \eta \tag{2.9}$$

where, for a finite lattice of $n = m \times m$ points:

- $p^*_{nx1}, S_{nx1}, \eta_{nx1}$

- $A_{nxn} \rightarrow$ Matrix with diagonal elements $-\tau\left(\frac{u}{\Delta x} + \frac{v}{\Delta y}\right)$

$$A = -\tau\left(\frac{u}{\Delta x} + \frac{v}{\Delta y}\right)I_n$$

- $B_{nxn} \rightarrow$ is a matrix which elements are the parameters that propagate the signal p^* along the W-E direction;
  Defining:
  $$\tilde{B}_{m \times m} = \frac{u\tau}{\Delta x}\begin{bmatrix} 0 & 0 & 0 & ... \\ 1 & 0 & 0 & ... \\ 0 & 1 & 0 & ... \\ ... & ... & ... & ... \end{bmatrix}$$
  $$B = \tilde{B}_{m \times m} \otimes I_m$$

- $C_{nxn} \rightarrow$ is a matrix which elements are the parameters that propagate the signal p^* along the S-N direction;

  Defining:
  $$\tilde{C}_{m \times m} = \frac{v\tau}{\Delta y}\begin{bmatrix} 0 & 1 & 0 & ... \\ 0 & 0 & 1 & ... \\ 0 & 0 & 0 & ... \\ ... & ... & ... & ... \end{bmatrix}$$
  $$C = I_m \otimes \tilde{C}_{m \times m}$$

Defining $W = A + B + C$, a more compact form of 2.9 is:

$$p^* = Wp^* + S + \eta \tag{2.10}$$

The model 2.10 identifies a *Simultaneous Autoregressive Model (SAR)* (e.g. Cressie, 1993, ch.6))

$$\mathbf{p}^* = \left(I - W\right)^{-1} S + \left(I - W\right)^{-1} \eta \quad (2.11)$$

which *first moment* is:

$$\mu = \left(I - W\right)^{-1} S \quad (2.12)$$

and for the properties of the *second moment* of a random vector η, the variance covariance matrix Σ for p^*:

$$\Sigma = Var\left[(I - W)^{-1}\eta\right] = \left(I - W\right)^{-1}\Lambda\left(I - W^T\right)^{-1} \quad (2.13)$$

Eq. 2.11 can be also be expressed as

$$(I - W)(p^* - \mu) = \eta \quad (2.14)$$

Setting $\Lambda = \sigma^2 I$ or, equivalently, $E[\eta\eta^T] = \sigma^2 I$ it is straightforward to define a multivariate Gaussian distribution for p^* over the spatial domain:

$$p^* \sim Gauss(\left(I - W\right)^{-1} S, \sigma^2 \left(I - W\right)^{-1} \left(I - W^T\right)^{-1}) \quad (2.15)$$

and the *log-likelihood function* as well:

$$-\frac{1}{2}ln2\pi\sigma^2 + \frac{1}{2}ln|Z| - \frac{1}{2\sigma^2}(p^* - \mu)Z(p^* - \mu) \quad (2.16)$$

where $Z = (I - W^T)(I - W)$ and $|Z|$ indicates the determinant of Z.

In applications, the horizontal wind components u and v as well as the spatial lags Δx and Δy are known, and τ and σ^2 are instead unknown parameters that can be estimated from observed data as shown in section 2.2.2.

Properties

In his founding work, Whittle, 1954 was the first who suggested a spatial autoregressive model to capture autocorrelation in the 2-D space based on the the definition of the *joint probability* of the process on the selected points of the space domain. His approach is in opposition to the one of Besag, 1974, based instead on the definition of *conditional probabilities* which posed the basis to what is referred to as Conditional Autoregressive Model (CAR). The distinction between the two, which is originally due to Brook, 1964, arises from the fact that, as opposed to time series where the definition of the *joint probability* and the *conditional probability* are equivalent, the two approaches lead to different results in the space domain.

The matrix W is interpreted as the *spatial-dependence matrix*: intuitively, taken two points on the lattice $(s_i), (s_j)$, $p^*(s_i)$ depends positively on $p^*(s_j)$ if $w_{ij} > 0$ while $w_{ij} = 0$ denotes independence. In other words, for the generic s_i site, the matrix W shows if the site s_j contributes to the value in s_i, by means of a corresponding non-zero entry in the ij cell.

Necessary and sufficient condition for the existence of the SAR model is $(I - W)$ be non-singular. A proof is found in Ripley, 2005. If $(I - W)$ is nonsingular, the covariance matrix Σ is strictly positive definite as required for a covariance matrix to be valid. For any specification of the structure of $(I - W)$ there will be a set of parameter values in W that make the matrix $(I - W)$ itself singular. These values depend on both the sample size (or size of the lattice) n, as well as the indexes of the non-zero elements in W (Anselin and Florax, 2012). That set of parameter values can be found solving the polynomial equation

$$|I - W| = 0 \tag{2.17}$$

for which there will be in general $k < n$ distinct roots that will depend on both n and the structure of W. Finding the set of values in W that makes the model non-valid in every circumstance it is applied is of crucial importance.

Useful properties on the parameters of SAR models are known and can be exploited when the structure of the matrix $(I - W)$ is of the kind $(I - \rho \tilde{W})$ where $\rho \in R$ is a parameter that can actually be interpreted as an autoregressive parameter. A common structure for $\tilde{W}$ is the so called *proximity matrix*, in which $\tilde{W}$ has unit entries in a cell when the two correspondent points are neighbors and 0 otherwise. Another widely used structure for $\tilde{W}$ is the so called *row stochastic matrix*, which elements in each row sum to 1. Good properties are known if $(I - W)$ has one of these two structures and especially when $\rho < 1$, and in those cases intervals of ρ for which $(I - \rho \tilde{W})$ is not singular are easy to find. The reader is referred to LeSage, 2008 for a comprehensive presentation. It is clear that the structure of

the SAR model derived in this chapter is more sophisticated and requires a further manipulation in order to make all the elements clear.

Reconsidering the elements present in the matrix $(I - W)$, its determinant depends on the position of the non-zero entries in W, the size n of the lattice, and the values of $\tau, u, v, \Delta x, \Delta y$. In applications, for a given problem, the quantities $u, v, \Delta x, \Delta y$ are known or fixed, while τ is the only parameter on which we can set constraints in order to $(I - W)$ be non singular. Besides the mathematical requirements, one should also remember that τ has a physical meaning, being a time. The constraints must therefore also respect his physical domain. As an example, in which a southwesterly flow is assumed, the case of a 2×2 lattice is considered here, setting $u = 12 m/s, v = 9 m/s, \Delta x = \Delta y = 1000 m$.

$(I - W)$ reads :

$$
\begin{bmatrix}
1 + \tau\left(\frac{u}{\Delta x} + \frac{v}{\Delta y}\right) & -\frac{\tau u}{\Delta x} & 0 & 0 \\
0 & 1 + \tau\left(\frac{u}{\Delta x} + \frac{v}{\Delta y}\right) & -\frac{\tau u}{\Delta x} & 0 \\
-\frac{u\tau}{\Delta x} & 0 & 1 + \tau\left(\frac{\tau u}{\Delta x} + \frac{\tau v}{\Delta y}\right) & -\frac{u}{\Delta x} \\
0 & -\frac{v\tau}{\Delta y} & 0 & 1 + \tau\left(\frac{u}{\Delta x} + \frac{v}{\Delta y}\right)
\end{bmatrix}
$$

and hence:

$$
\begin{bmatrix}
1 + \tau * 0.021 & -\tau * 0.012 & 0 & 0 \\
0 & 1 + \tau * 0.021 & -\tau * 0.012 & 0 \\
-\tau * 0.009 & 0 & 1 + \tau * 0.021 & -\tau * 0.012 \\
0 & -\tau * 0.009 & 0 & 1 + \tau * 0.021
\end{bmatrix}
$$

with the following *characteristic polynomial* :

$$(1.4\, e^- 7)\tau^4 + (3.4 e^- 5)\tau^3 + 0.0026\tau^2 + 0.084\tau + 1$$

which is greater than 0 for every positive τ and so also for any value in its feasible range $\tau \in [400, 4000]$. When both u, v are positive, any change in the length of the spatial lags $\Delta x, \Delta y$ or in the velocities u, v or again in the size n of the lattice would not modify that positive values of τ are not roots of the *characteristic polynomial*.

W does not have to be symmetric, meaning $w_{ij} = w_{ji}$ is not a requirement. Most authors state the assumption $w_{ii} = 0, i = 1, ...n$, although there is not necessarily any theoretical constraint on the diagonal of W to be a 0-vector for model validity, as long as $(I - W)$ is not singular. In the case treated here $diag(W) = diag(A)$ and thus has elements $-\tau\left(\frac{u}{\Delta x} + \frac{v}{\Delta y}\right)$

The covariance matrix Σ is a full matrix: each point on the lattice is correlated to all the other points at an extent that decreases with the decreasing order of proximity, but that cannot be directly established simply on the basis of the sites pairwise spatial distance. Moreover, despite the introduction of an homoskedastic error term $\eta \sim Gauss(0, \sigma^2 I)$, the autocorrelation in p^* induce heteroskedasticity it the latter, being the elements on the main diagonal of Σ not all equal.

If we consider the dependence structure between p^* and the stochastic noise η, it is easily shown that

$$cov(p^*, \epsilon) = E[p^* \eta^T] = \Lambda (I - W^T)^{-1} \tag{2.18}$$

which is not diagonal, meaning that the error η and p^* are not stochastically independent.

The two models are here contrasted using the same notation, to make comparison between the two easier and to present how they can be conceptually and theoretically linked under certain circumstances. Following Cressie, 1993, Ripley, 2005 and Anselin and Florax, 2012 :

SAR model:

$$(I - W)(p^* - \mu) = \eta; \qquad p^* \sim Gauss(\mu, (I - W^T)^{-1} \Lambda (I - W^T)^{-1}) \tag{2.19}$$

or marginally for one generic single point

$$p^*(s_i) = \mu_i + \sum_{i=1}^{n} w_{ij}(p^*(s_j) - \mu_j) + \eta_i \tag{2.20}$$

CAR model:

$$(I - C)(p^* - \mu) = \nu; \qquad p^* \sim Gauss(\mu, (I - C^T)^{-1} M) \tag{2.21}$$

where $M = diag(\tau_1, ..., \tau_i, ..., \tau_n)$ contains in the main diagonal the conditional variances associated to sites $s_1, ..., s_n$.

The *log - likelihood* has the same structure of the one of the SAR model, namely

$$-\frac{1}{2} ln 2\pi\sigma^2 + \frac{1}{2} ln|Z| - \frac{1}{2\sigma^2}(p^* - \mu)Z(p^* - \mu) \tag{2.22}$$

but where now $Z = (I - C)$.

Looking at the expression for one generic single point:

$$p^*(s_i) = \mu_i + \sum_{i=1}^{n} c_{ij}(p^*(s_j) - \mu_j) + \nu_i \tag{2.23}$$

For this model $E[p^*\nu^T] = M$, and being M diagonal it means the stochastic error term is not stochastically dependent on p^*.

where setting τ_i and τ_j the conditional variance for the points s_i and s_j respectively, the elements of the matrix C should be such that $\tau_i c_{ij} = \tau_j c_{ji}$ and $c_{ij} = 0$ if sites s_i and s_j are independent.

This is equivalent to require the matrix $(I - C)$ be non singular and the covariance matrix $(I - C)^{-1}M$ be symmetric and positive-definite.

The two approaches are not in general equivalent. They lead to the same model if and only if $(I - C)^{-1} = (I - W)^{-1}\Lambda(I - W^T)^{-1}$, given a proper definition of a mean μ such that it makes no difference in the choice between the two models Cressie, 1993. When these last requirements are met, any SAR model can be represented as a CAR model, but the opposite is not always necessarily the case. From the previous equivalence between the covariance matrices, it is clear that even in case the two approaches led to the same model, C would differ from W. More precisely, any SAR model is a CAR model with $C = W + W^T - W^TW$. In those cases when a CAR model can be seen as a SAR model, the reverse relation between the two spatial dependence matrices W and C takes a less natural form, since $W = I - L^T$ where LL^T is the Cholesky decomposition of $(I - C)$ Ripley, 2005.

Although the conditional dependence specified by a CAR model is often more natural than the one created by a SAR model and although working with a CAR model poses fewer computational issues, the SAR model of Eq. 2.11 does not allowed a straightforward specification of an equivalent CAR model. Indeed, although the CAR model theoretically derived would still be a valid model, when the means of the two models are contrasted, the choice between $(I - W)^{-1}S$ and $(L^T)^{-1}S$ is not equivalent. Choosing the second representation of the mean would introduce artificial symmetries in the mean field of p^* that again would be unphysical. The following Fig. 2.1 gives a visual insight of the issue: for the idealized terrain displayed in the left panel and wind components $u = 12.65m/s, v = 9.06m/s$, the mean field of p^* is displayed as derived by the SAR model (central panel) and the CAR model (right panel):

The SAR model captures well as expected the advection toward the North-East corner of the cloud water as well as positive values of the mean of p^* in the upslope
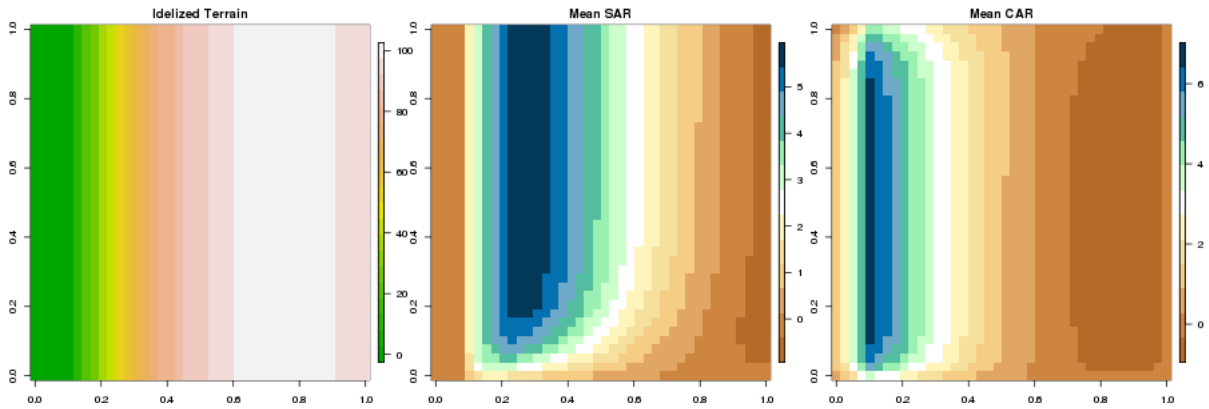
Fig. 2.1: Comparison of the mean of a SAR model and a CAR model on an idealized terrain.

region. The CAR model, instead, introduces an unphysical symmetry as well as not expected regions with positive values in the mean of p^*.

In the SAR model the dependence at the edges of the lattice, being simultaneously specified for every site, does not pose any issue. In contrast, the CAR model suffers from the so-called edge effect, that requires a more sophisticated determination of the dependence structure for the sites at the edges of the lattice.

Comparison with Smith's Model

The statistical model derived in Eq. 2.11 is not just a discretized version of Smith's model. Firs of all it allows to estimate from data the unknown value of the parameter τ, instead of requiring tuning procedures.

The statistical model is more flexible: it basically controls for two elements, namely the production of cloud water and the progation of the signal across the space domain. For the first element, it allows to use a spatially varying values of the large-scale variables used to estimate the source S for each site, instead of a single value for the entire domain as in the deterministic model. Regarding how the signal is propagated, the statistical in Eq. 2.11 could be straightforwardly extended allowing the intensity of the horizontal wind used for building W to vary in space, again in opposition to the deterministic model. Moreover, once the structure of the statistical model has been recognized, many modifications and extensions are allowed that will let the model valid. One possible extension is the one presented in section 2.3

2.2.2 Estimation

In the case of gaussian data, since p^* is stochastically dependent on η, using least squares for estimates would not be consistent Whittle, 1954. Insightful discussions

on the methods for parameter estimation using approaches based on likelihood can be found in Cressie, 1993, Ripley, 2005 among others. LeSage, 2008 presented both likelihood estimation (ch.2) and methods based on the Bayesian approach (ch.5).

Records of observed precipitation take values in R^+, and aiming at decomposing the process into two steps "occurrence" and "intensity" of precipitation, a dichotomous variable would indicate the presence, 1 or the absence, 0 of recorded precipitation. In both cases the methods used for gaussian data are not directly applicable and diverse approaches must be followed.

The SAR Probit Model for Precipitation Occurrence

When modeling precipitation occurrence the dependent variables for which we can get observations at the ground are binary variables with $\{0, 1\}$ responses which reflect absence and presence of precipitation respectively. Equal values are more likely observed at close location if a spatial dependence does exist. In literature, models built with the aim of modeling binary data are referred to as *Limited dependent variables spatial models*, in which one identifies adequate linkages between the binary observed response and a modeled latent unobserved process revealed by the binary variable.

The *Latent Response Model*, (Verbeek, 2008, p.180) is a representation commonly used for this scope that makes use of the concept of a *latent process*. Following this approach when modeling binary dependent variables, the observed dichotomous response $y \in \{0, 1\}$ is treated as indicators of the latent, continuous unobserved process. Among *latent response models*, the *Probit Model* constitute one possible approach. In particular, in the case treated in this thesis, once the SAR model has been derived (2.10) for the latent, unobservable process p^*, the probit model assesses an adequate linkage between p^* and the observations y, defining what is usually referred to as a *SAR probit model* (e.g LeSage, 2008). More precisely, the latter model assumes the latent unobserved process p^* follows a normal distribution such that the observed $y \in \{0, 1\}$ values are reflected in:

$$
\begin{aligned}
y_i &= 1, \; if \; p_i^* > 0 \\
y_i &= 0, \; otherwise
\end{aligned}
\tag{2.24}
$$

McMillen, 1992 proposed an EM algorithm (Dempster et al., 1977 is a classical reference) for a Probit variant of the SAR model that produces consistent maximum likelihood estimates. In each iteration t until convergence, the method estimates: in the E-step the vector of the expected values of p^* given the observed data y and which depends on the estimated value of τ obtained in the previous iteration,

namely $\hat{E}[p^*|y, \hat{\tau}^{t-1}]$; then, in the M step, those estimates are directly introduced in a *log-likelihood* function for τ, and an estimate at iteration t for the latter is obtained maximizing that function. The value obtain at this stage will be used in the E-step of the successive iteration $t + 1$. The method produces biased yet consistent estimates of the covariance matrix (McMillen, 1992).

LeSage, 2000 proposed a method based on Markov Chain Monte Calro (MCMC) samples that overcomes some of the limitations in McMillen, 1992 moving from the works by Albert and Chib, 1993 for non-spatial probit models and LeSage 1997. Another option is to consider the so-called Tobit (or censored regression) model variant of the SAR model (e.g LeSage, 2008). An historic and conceptual review of these approaches is found in LeSage, 2008, while Fleming, 2004 gave a more extensive discussion contrasting alternative approaches to estimating parameters for SAR Probit models and other Limited dependent variable spatial models.

The parameter estimation of the SAR probit model (2.10, 2.24) for precipitation occurrence is done here after adapting the method by LeSage, 2000 and nesting the one by Li and Ghosh, 2015 as it will be described below. The adaptation makes possible to account for the different model specification of Eq. 2.11, and the different unknown parameters involved.

For an introduction to Bayesian methods for spatial data the reader is referred to Banerjee et al., 2014. In the Bayesian approach followed here, given the observed vector with binary elements y, the $n \times 1$ elements of the unobserved p^* are considered parameters to be estimated as well through the procedure. Setting *a priori* distributions on τ, σ^2, and after the estimation of p^* it is in principle possible to proceed and estimate the remaining model parameters τ, σ^2, sampling in particular from their posterior distribution $(\tau, \sigma^2|p^*, y)$. Hereafter the parameter σ^2 is considered fixed and known.

What makes possible the use of this approach is a result found by Albert and Chib, 1993: they argued that in the case the vector of the latent process p^* were known, y would be known as well. With an abuse of language, it means the information carried by p^* contains all the information carried by y (the opposite untrue), and that allowed them to conclude:

$$p(\tau, \sigma^2|y*) = p(\tau, \sigma^2|y*, y) \tag{2.25}$$

In other words, treating p^* as an additional set of parameters to be estimated, the conditional posterior distribution for the parameters τ, σ^2 conditioning on both p^* and y takes the form of a Bayesian regression problem for a continuous dependent variable instead of a problem involving the binary-value vector y.

To estimate the vector of n parameters $p^*_i, i = 1, ..., n$, the estimation via MCMC sampling requires to draw, at any pass, sequential samples from any individual p^*_i from its conditional distributions $p(p^*_i | p^*_{-i})$, thus conditional on all the other elements in the vector p^*_{-i} excluding the $i - th$.

Through a large number of passes, this sampling method produces a sequence of draws for the parameters that converges to the joint posterior distribution $p(p^*, \tau | y)$ or eventually $p(p^*, \tau, \sigma^2 | y)$.

Albert and Chib, 1993 considered the case of non-spatial regression model (independent observations) for binary observed data y and found a conditional distribution for the individual p^* that takes the form of an *univariate truncated normal* distributions, namely a normal distribution whit bounded values. (lower, upper or both). See Li and Ghosh, 2015 for a clear description and properties of the truncated normal distribution and sampling methods.

Dealing with a SAR model for spatially dependent observations, the previous approach leads to a *multivariate truncated normal* (MVTN) distribution (e.g Li and Ghosh, 2015 for properties of the MVTN) for the latent p^* parameters from which one needs to sample. The condition 2.25 still holds: while the individual conditional posterior distribution $p(\tau | \sigma^2, y*)$ is as in the case of a continuous dependent variable, drawing from the $n-$variate truncated normal for p^* is not straightforward. Geweke, 1991 pointed out that this cannot be done for the case of a truncated multivariate distribution. That is, the individual elements from a vector such as $y*$ cannot be obtained by sampling from a sequence of univariate truncated normal distributions as in the case of independent observations.

An MCMC sampler for the SAR probit model

The scheme of the MCMC sampler adapted from LeSage, 2000 is the following: the scheme requires to sample sequentially from the conditional posterior distributions for τ, p^*. Within this sequence, it is also needed to sample a set of n values to fill-in the vector p^*.

A *uniform distribution* was chosen for τ in its feasible range $[400, 4000]$: $p(\tau) = \frac{1}{4000-400}$. The algorithm is then composed of the following steps:

1. sampling τ from its individual conditional posterior distribution $p(\tau | p^*)$. Making explicit $W = \tau \tilde{W}$, the following expression was found:

$$p(\tau | p^*) \propto |I_n - \tau \tilde{W}| exp\big(-\frac{1}{2\sigma^2}[(I_n - \tau \tilde{W})p^* - S]^T[(I_n - \tau \tilde{W})p^* - S]\big) \quad (2.26)$$

where $|I_n - \tau \tilde{W}|$ is the determinant of $I_n - \tau \tilde{W}$. Samplings was then accomplished using a *Metropolis-Hastings algorithm*.

2. Sampling each p^*_i from its conditional distribution. Issues on this second sampling procedure were pointed out before. Finding methodologies to draw from this distribution is an active branch of research. One well known method is the one proposed by Geweke, 1991, but here the more efficient adaptive Gibbs sampling introduced by Li and Ghosh, 2015 was used. The method allows to derive the conditional univariate truncated normal distribution for a transformed variable of p^* for each site. The distribution is such that it is easy to sample from it. The truncation boundaries of the transformed variable are determined from the values of the observations y. The method is also efficient because it adopts 3 different sampling algorithms depending on the values of the boundaries, optimizing the acceptance rate of the algorithm.

Each pass of the MCMC sampler constitutes only a single sample of τ and one single sample for the entire vector p^*. A large number of passes must be done to produce a large sample of draws from the joint posterior distribution of the model parameters. Once the sample of draws is obtained, it can be used to construct parameter estimates based on statistics of the posterior distributions.

SAR model for precipitation intensity

Precipitation can be considered a mixture of discrete and continuous distributions, since precipitation is a continuous variable with exact zeros if no precipitation is recorded Dunn, 2004. Following a *chain-dependent* stochastic approach (Katz, 1977), the discrete and the continuous part are modeled separately in two-stages. In literature many probability distributions for precipitation intensity has been suggested. The reader is referred to Wilks, 2011 and references therein for a review and to and Wilks and Wilby, 1999 for a comprehensive historical picture.

In other works, it is assumed both precipitation occurrence and itensity can be modeled using the same latent Gaussian process. Various known transformation functions have been suggested to transform the Gaussian values to the desired intensities. For example, Katz and Parlange, 1995 used a power transform with exponent smaller than 1, Allcroft and Glasbey, 2003 used a quadratic power function, Sigrist et al., 2012 used a power function, and Allard and Bourotte, 2015 used a power-exponential function . Kleiber et al., 2012 applied a two-part transformation function, with one part being the standard normal distribution and the second part the inverse of a gamma distribution.

A discussion on the choice or definition of a suitable probability density function for precipitation intensity or transform of a gaussian latent process is beyond the scope

of this work. A general structure will be kept, choosing for the case studies presented in chapter 3 a power transform based on the approach in Katz and Parlange, 1995. The intensities are found applying a power transform of order $k, 0 < k < 1$ to the latent gaussian field p^* on "wet" sites, when $p^* > 0$. k is chosen as the value that makes the $95th$ percentile of both the distributions be equal. Namely, letting y be the observed precipitation on "wet" sites:

$$k : (p^*)^k_{[.95]} = y_{[.95]} \tag{2.27}$$

This choice lets the spatial dependence be entirely determined in the latent process p^* and in the estimate of the $95th$ percentile treats the observed values as observations from i.i.d variables.

2.2.3 Prediction

Once the parameters τ and eventually σ^2 are estimated as described in the previous section, it is possible to predict the values of precipitation occurrence and amount on unsampled (not observed) sites.

This can be achieved determining the values of the source function S, of the mean $\mu = (I - W)^{-1}S$ and of the covariance matrix $\Sigma = \left(I - W\right)^{-1} \sigma^2 I \left(I - W^T\right)^{-1}$ for the unsampled sites and generating the field p^* from a multivariate gaussian distribution and consequently occurrence y and intensity x on those spatial points.

In order to calculate μ and Σ all the values required are now known.

At this stage it is also possible to modify the grid, namely modifying the values of $\Delta x, \Delta y$. It is instead not possible to directly modify the components u, v of the wind speed. That would require recomputing the source function of condensation S, as well as getting a new estimate for τ, since that reformulation would imply a different spatial pattern of precipitation at the ground.

2.2.4 Case 2: Irregularly Spaced Precipitation Data

The model in 2.11 is defined for observed data on a regular lattice. Observations of precipitation are often recorded at irregularly spaced sites. To combine the advantages of the proposed scheme with the flexibility required in sample designs, an adaptation of the model formulation is required.

Although building the source function S is straightforward on unevenly spaced points, the major issue is faced in modeling the spatial dependence structure, and thus in defining the proximity matrix W. That might be handle taking advantage from the fact that matrix W contains the measures Δ_x, Δ_y of the distance between the pairs of points. From a theoretical point of view, the values of elements Δ_x, Δ_y, present at the denominator of each non-zero entry of W, can vary among pairs of points, and so across the entries of W, as long as the resulting matrix $(I - W)$ is non-singular for model validity. The issue is thus restricted to the choice of neighbors for each site s_i among all the other spatial points $s_{j,j \neq i}$, and in setting the proper corresponding entry (i, j) in matrix $W = A + B + C$, by specifying the correct pairwise distances $(\Delta_x(i, j), \Delta_y(i, j))$.

White and Ghosh, 2009 proposed a neighbor function for defining a CAR model on an irregular lattice. Here a similar approach is followed and a new neighbor function is defined which accomodates to the peculiarities of model 2.11.

For the identification of a neighbor, a maximum distance d_{max} is set. The distance should be set small enough for preventing far points to have an influence and at the same time big enough to ensure matrix W has at least one non-zero entry in each row in addition to the elements on the main diagonal. Thus, parameter d_{max} ensures the resulting matrix $(I - W)$ is sparse. Moreover, for site s_i, only sites in the upwind sector with respect to s_i can become its neighboring points. Thus, in case of a Southwesterly wind, neighbors of point s_i must be searched in the sector South-West of it. A second parameter d_l might be necessary in some case to represent an intermediate distance. In particular, pairs of points closer that d_l are considered as they were neighbors on a regular lattice, and thus set as distant as at the artificial distance d_l itself. This artificial spacing avoids the distances Δ_x and Δ_y to be too small, and thus matrix W and the covariance matrix to have single entries with very high values that might corrupt the estimates. Moreover, d_l and d_{max} could serve in ensuring matrix $(I - W)$ is non-singular. Matrices A,B,C necessary to build eq. 2.9, are consequently built as hereafter explained, taking again as example a Southwesterly wind.

Building matrix B: matrix B explains the signal propagation in the longitudinal direction. The distances $\Delta_x(i, j)$ at the denominator of the (i, j) entries of matrix B are determined in the following way:

$$\Delta_x(i, j) = \begin{cases} d_{x;l} & \text{if } 0 < d_{ij} < d_l \\ d_{x;ij} & if \ d_l < d_{ij} < d_{max} \\ \infty & otherwise \end{cases} \quad (2.28)$$

where d_{ij} is the distance between the sites s_i, s_j, $d_{x;ij}$ is the distance in the longitudinal direction between sites s_i and s_j and $d_{x;l}$ is the longitudinal component of the artificial minimum distance d_l. The case $\Delta_x(i,j) = \infty$ is equivalent to not considering the point s_j as neighbor and influential of s_i. Thus, a $n \times n$ matrix $\tilde{B}$ is created by inserting the elements $\frac{1}{\Delta_x(i,j)}$ in the corresponding (i,j) entries. Finally, matrix B is:

$$B = u\tau\tilde{B} \tag{2.29}$$

Building matrix C: matrix C explains the signal propagation in latitude. The distances $\Delta_y(i,j)$ at the denominator of the (i,j) entries of matrix C are determined in the following way:

$$\Delta_y(i,j) = \begin{cases} d_{y;l} & \text{if } 0 < d_{ij} < d_l \\ d_{y;ij} & if\ d_l < d_{ij} < d_{max} \\ \infty & otherwise \end{cases} \tag{2.30}$$

where the elements are interpreted similarly as those presented for matrix B. Thus, a $n \times n$ matrix $\tilde{C}$ is created by inserting the elements $\frac{1}{\Delta_y(i,j)}$ in the corresponding (i,j) entries. Finally, matrix C is:

$$C = v\tau\tilde{C} \tag{2.31}$$

Building matrix A: in model 2.11 defined on the regular lattice, matrix A was: $A = -\tau\left(\frac{u}{\Delta x} + \frac{v}{\Delta y}\right)I_n$. Now Δ_x, Δ_y must be substituted by the average values $\Delta_x = \overline{\Delta_x}, \Delta_y = \overline{\Delta_y}$ computed among the values used to build matrices B and C respectively. Or, if values $d_{x;l}, d_{y;l}$ were set, then those would be the new Δ_x, Δ_y to use in matrix A. Thus, after defining appropriate Δ_x, Δ_y, matrix A is:

$$A = -\tau\left(\frac{u}{\Delta x} + \frac{v}{\Delta y}\right)I_n \tag{2.32}$$

As an example, imagine to observe precipitation data on the 20 spatial points depicted in fig. 2.2:

and to build matrix B for a Southwesterly wind. Let us set $d_{max} = 2.5$. For point (1), we look for suitable neighbors South West of the point itself. Neighbor point of (1) is (14), with $d_{x;ij} = 2$. Thus, the first row of B will have all zero entries, but $1/2$ in the 14th column. Neighbors of point (2) are (7) and (16), with distances $d_{x;ij} = 1$ and $d_{x;ij} = 2$ respectively. Thus the second row of matrix B will have entries 1 and $1/2$ in the corresponding 7th and 16th columns.
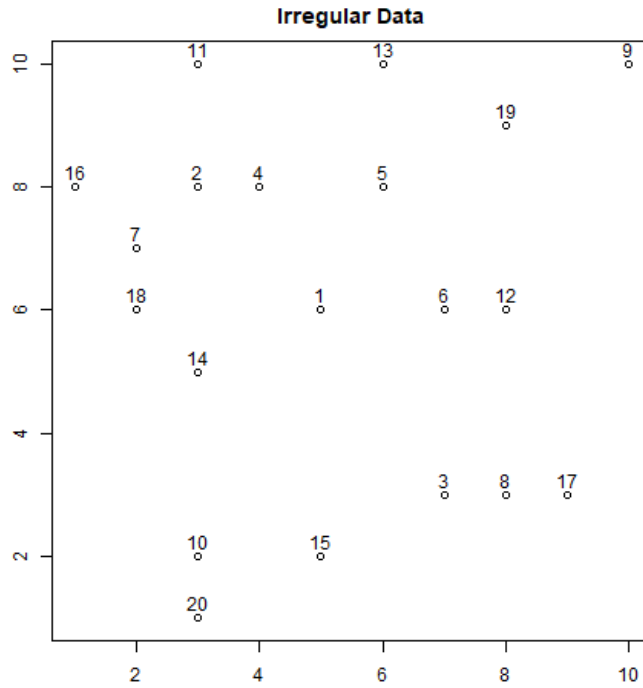
Fig. 2.2: Example of 20 irregularly spaced points over

The same procedure is used to build matrix C.

2.3 Inclusion of large-scale precipitation

The statistical model defined in Eq. 2.11 computes and propagates cloud water arising from up-slope orographic mechanisms. Large scale precipitation fraction from frontal systems or wide convective cores is not directly taken into account.

To allow the inclusion of these large scale components of precipitation, two extensions of the model are here suggested. If knowledge of the large scale precipitation (LSP hereafter) on the study area is available (for example from climate/forecast model or reanalysis), it is possible to add its contribution, modifying 2.10 as follows:

$$p^* = Wp^* + \alpha * LSP + S + \eta \tag{2.33}$$

with mean $\mu = (I - W)^{-1}(\alpha * LSP + S)$ and the same covariance matrix as before.

α is an unknown parameter that must be estimated from data. It modulates the direct impact of the large scale precipitation on the spatial distribution, in order

to avoid shocks that could arise introducing directly the value of the intensity of LSP obtained from climate models, and using instead the information in observed data to complement it. This choice also reflects that the orographic enhancement of preexisting storms vary strongly for frontal systems reaching their maturity before or after the passage over the mountain range. Corradini et al., 2000 suggested that the enhancement is of the order of 50% in prefrontal storms to 150% in postfronatl storms. The reader is referred to Houze, 2012 and Barros and Lettenmaier, 1994 for a deep review of the features of baroclinic disturbances and their interaction with mountains.

The second extension further adds a third unknown parameter as *weight factor* of the orographic source S. Letting the design matrix become the $n \times 2$ matrix $X = [S, LSP]$ and the new two-dimensional parameter vector $\lambda = (\lambda_1, \lambda_2)^T$, the extended model reads:'

$$p^* = (I - W)^{-1}(X\lambda) + (I - W)^{-1}\eta \tag{2.34}$$

with mean $\mu = (I-W)^{-1}(X\lambda) = (I-W)^{-1}(\lambda_1 S + \lambda_2 LSP)$ and same covariance matrix as before. λ_1, λ_2 will thus weight the small scale orographic (S) and large-scale (LSP) components to let total precipitation match observations. Here λ_1 modulates the effect of the orographic cloud source S. In particular, the Source function 2.3 is built under the assumption of nearly saturated atmosphere, an assumption that often leads to overestimate the actual condensation and precipitation rates. λ_1 then could act as correction of this strong assumption, reducing the rate of cloud production thanks to the information inferred from observed precipitation data used to estimate λ_1

2.3.1 Estimation

The estimation procedure exposed in the previous section must be modified to include new unknown parameters. Here the estimation procedure for model 2.34 is presented. Additional full conditionals must be found for $\lambda = (\lambda_1, \lambda_2)$, and the one for τ must be modified accordingly as well. Something to note is that the inclusion of an unknown parameter in the mean μ, will lead to an identification issue: in fact in the probit SAR model, different pairs of values of σ^2, α, would lead to the same value of the likelihood function of p^* (Geweke, 1993). So in this case the value of σ^2 must be fixed as a proper constant.

Recalling the procedure exposed in section 2.2.2, an uninformative *a priori* is set again for τ: $\tau \sim Unif(400, 6000)$. The upper bound was augmented to 6000s after noticing the natural behaviour of the chains. A bivariate gaussian distribution is

chosen for λ: $\lambda \sim N(0, T), T = 1e + 10I_2$, with zero mean, zero covariance and uninformative standard deviation.

The three main steps of the sampling scheme then become:

1 Sample each individual $p_i^*, i = 1, ..., n$ from its full conditional posterior distribution $p(p_i^* | p_{-i}^*, \tau, \lambda, y)$

2 Sample τ from $p(\tau | \lambda, ^*, y)$ via Metropolis-Hastings. Making explicit $W = \tau \tilde{W}$, the kernel of $p(\tau | \lambda, ^*, y)$ is:

$$p(\tau | p^*, \lambda, y) \propto |I_n - \tau \tilde{W}| exp\left(-\frac{1}{2\sigma^2}[(I_n - \tau \tilde{W})p^* - \lambda X]^T[(I_n - \tau \tilde{W})p^* - \lambda X]\right)$$
(2.35)

3 Sample λ from $p(\lambda | \tau, p^*, y)$: $p(\lambda | \tau, p^*, y) \sim N_{MV}(d^*, \sigma^2 f^*)$, a bivariate Gaussian distribution where

$$d^* = (X^T X - T^{-1})^{-1}(X^T(I_n - \tau \tilde{W})p^*)$$

$$f^* = (X^T X - T^{-1})^{-1}$$

which is characterized by small negative correlation between the two coefficients λ_1, λ_2.

Results

<div style="text-align: right; font-size: 2em;">3</div>

3.1 Idealized Experiments

This section reports a set of idealized experiments carried out simulating from the statistical model in eq. 2.11, prescribing some values of the unknown parameters. Four different idealized terrains are built and used in four experiments. Each experiment is based on the same set of 3 different environmental conditions, making possible to confront the behavior of the statistical model in a total of 12 comparable experiments.

The idealized ridges are chosen in order to identify a few yet clear features of a real topography: N-S oriented long ridges allow a straightforward comparison of the effect of different mountain slopes on precipitation patterns in a setting close to some of the main real mountain ridges (e.g. Cascades (North-West USA), Sierra Nevada (CA, USA), Andes (South America)); an isolated circular Gaussian hill constitutes a reference experiment resembling isolated high peaks; the last idealized terrain used is composed of N-S oriented multiple ridges. All the idealized terrains have size and proportions reasonably resembling a real terrain.

The atmospheric variables used to build the quantities in the source function (eq. 2.3) as well as the horizontal components of the wind velocity are taken from a sample day in winter from the Era-interim dataset (Dee et al., 2011) for a box over the Coastal Range and Sierra Nevada in California, US. This allows testing the model using realistic variables consistent with each other. The vertical levels used for all the variables are in the interval 850hPa - 700 hPa, motivated by the results in Neiman et al., 2002, who found the highest correlation between the flow velocity and rainfall over the California Coastal Range is found at approximately the same altitude of the peak of mountain ranges. The value of the saturation water density is set to $\rho_{sref} = 0.009[Kg/(m^2)]$

Furthermore, in each experiment and sub-experiments, the time-delay parameter τ is set to 8 different values, namely $(1, 400, 800, 1000, 1500, 2000, 4000)$ seconds. $\tau = 1$ approximates a simple upslope model where water falls immediately to the ground after condensation.

All the simulations from the statistical model are realized using the same seed and the same value of σ^2 for the covariance matrix $\Lambda = \sigma^2 I$ of the noise term $\eta \sim N(0, \Lambda)$.

North-South Oriented Mountain Range

An idealized mountain range was built using N-S oriented long mountain range. The shape of the East-West section is given by a Gamma probability density function, as shown in Fig. 3.1 left panel. The section is identical in the entire North-South direction and the altitude h is built as:

$$h(x) = A \cdot \frac{1}{\Gamma(k)\theta^k} \cdot x^{(k-1)} \cdot exp(-\frac{x}{\theta})$$

(3.1)

where here the values $A = 40, k = 2, \theta = 25$ were used.
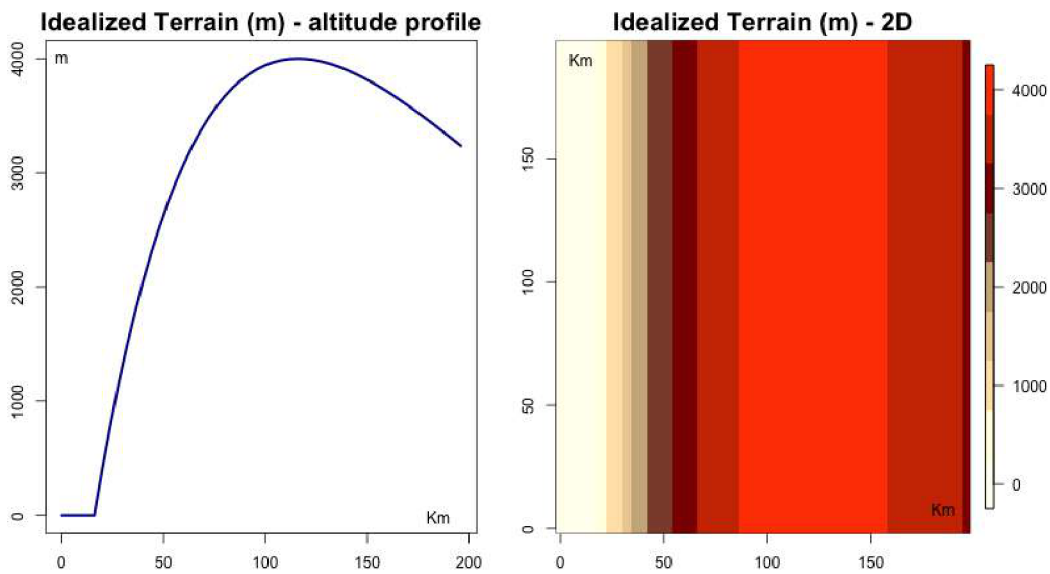
Fig. 3.1: Left: W-E section of the idealized terrain; Right: 2D representation of the terrain

The range is asymmetric in the longitudinal dimension, the highest elevation is 4000m and the spatial domain is squared with side 200 Km. The lattice is composed of 50×50 points separated by $\Delta_x = \Delta_y = 4000m$. The experiment is composed of three sub-experiments (A,B,C), showing how the simulation from the model changes by prescribing the following three environmental conditions:

- A: horizontal component of the wind, homogeneous over the area: $u = 10m/s$, $v = 10m/s, RH = 35\%$

- B: horizontal component of the wind, homogeneous over the area: $u = 15m/s$, $v = 10m/s, RH = 35\%$

- C: horizontal component of the wind, homogeneous over the area: $u = 10m/s$, $v = 15m/s, RH = 35\%$

Sub-Experiment - A Fig. 3.2 left panel depicts the source function $S(x,y)$ as in eq. 2.3 where the quantities are derived for $u = 10m/s$, $v = 10m/s$, $RH = 35\%$

Fig. 3.40 shows the corresponding simulations from the model letting τ vary.

Sub-Experiment - B Fig. 3.2 central panel depicts the source function $S(x,y)$ where the quantities are derived for $u = 15m/s$, $v = 10m/s$, $RH = 35\%$. The simulations relative to this experiment are found in Fig. 3.41

Sub-Experiment - C Fig. 3.2 right panel depicts the source function $S(x,y)$ where the quantities are derived for $u = 10m/s$, $v = 15m/s$, $RH = 35\%$. The simulations relative to this experiment are found in Fig. 3.42

In all the three simulations the maximum of precipitation is found windward before the crest. The higher the value of τ the further the the condensed water falls on the leeside. The advection is also visible in the bottom edge of the lattice where the movement of the precipitation pattern northward is visible as the value of τ increases. For a given source of condensed water, the intensity of precipitation in each location decreases as τ increases. The level of spatial dependence among the points increases with the value of τ, making the propagation of the signal less oscillating.

Modifying the horizontal components of the wind, the direction and intensity of the advection is modified as well. Because of how the source function of condensation is built, increasing the horizontal wind speed also the vertical velocity is increased end hence the rate of condensation.
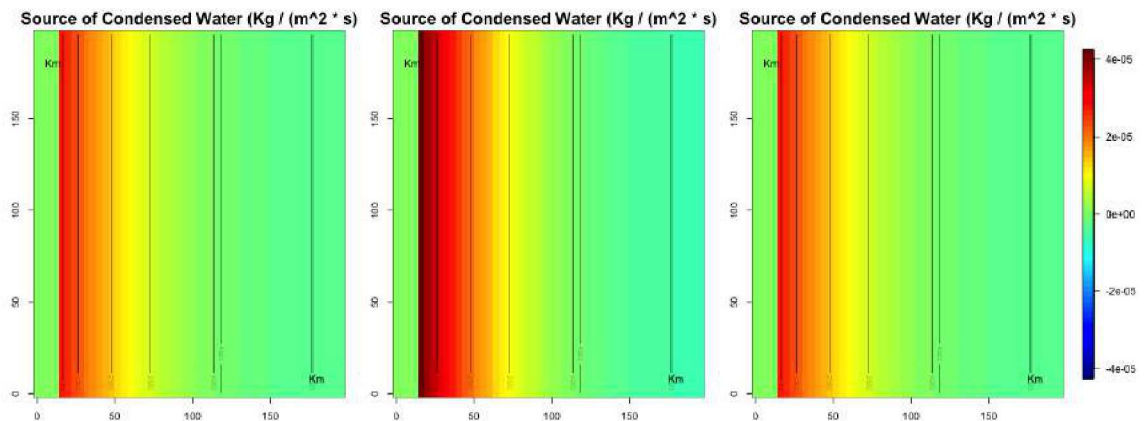
Fig. 3.2: Source function related to the three sub-experiments: A) left panel; B) central panel; C) right panel

The same 3 sub-experiments are repeated varying the shape of the mountain flipping along the north-south axes the ridge (Fig. 3.3). The slope is less steep upwind and more steep downwind. This shape creates less amount of condensed water upwind of the peak and a sharper shadow effect.
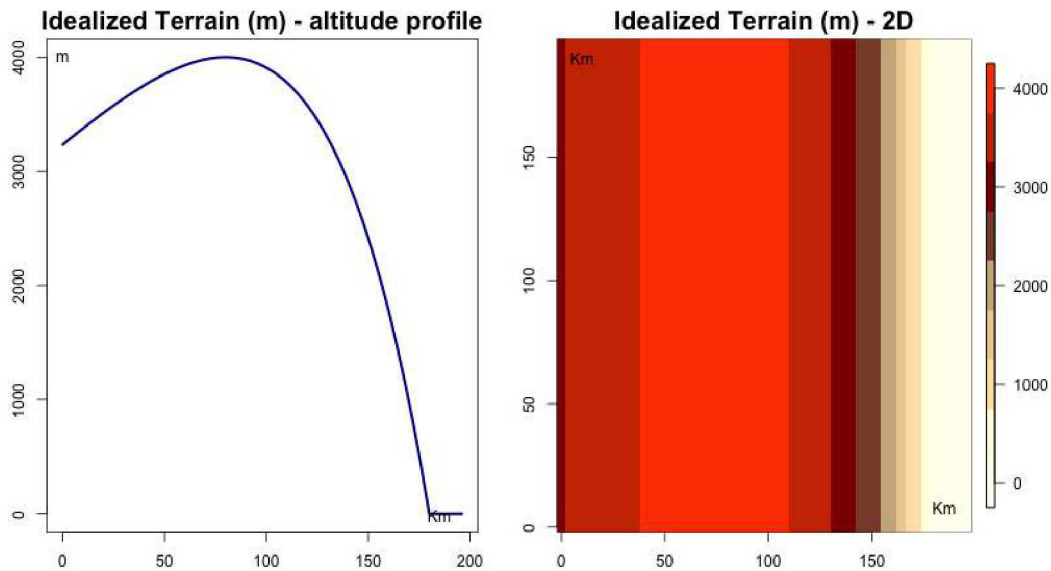
Fig. 3.3: Left: W-E section of the idealized terrain; Right: 2D representation of the terrain

Sub-Experiment - A Fig. 3.4 left panel depict the source function $S(x, y)$ induced by the new terrain and where the quantities are derived for $u = 10m/s$, $v = 10m/s$, $RH = 35\%$

The corresponding simulations form the model are presented in Fig. 3.43

Sub-Experiment - B Fig. 3.4 central panel depict the source function $S(x, y)$ where the quantities are derived for $u = 15m/s$, $v = 10m/s$, $RH = 35\%$

The corresponding simulations form the model are presented in Fig. 3.44

Sub-Experiment - C Fig. 3.4 right panel depicts the source function $S(x, y)$ where the quantities are derived for $u = 10m/s$, $v = 15m/s$, $RH = 35\%$

The corresponding simulations form the model are presented in Fig. 3.45
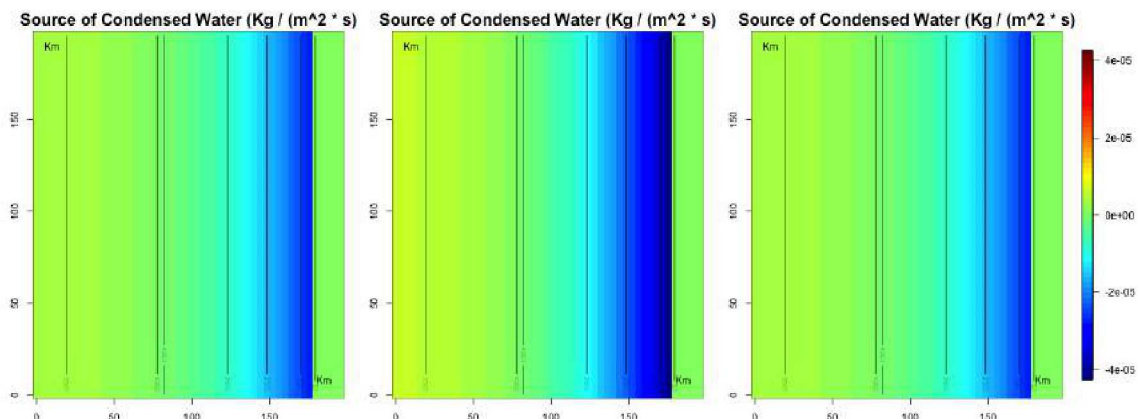
With respect to the previous experiment, the less steep slope in the windward side and the steeper one on the leeside produces less amount of precipitation and a sharper shadow effect. Being the source of condensation sensibly smaller than in the previous experiment, the overall amount of precipitation at the ground is greatly reduced. Because of the decision of keeping fixed the value of the parameter σ^2, the signal to noise ratio in this experiment is smaller.

Isolated Circular Gaussian Hill

The second experiment is carried out using an isolated gaussian hill as idealized terrain (Fig. 3.5) and the same environmental conditions as in the previous experiment. The terrain elevation is derived as:

$$h(x,y) = A \cdot exp\left\{ -\left[\left(\frac{x}{a_x}\right)^2 + \left(\frac{y}{a_y}\right)^2 \right] \right\} \quad (3.2)$$

where a_x and a_y are the E-W and N-S width respectively.
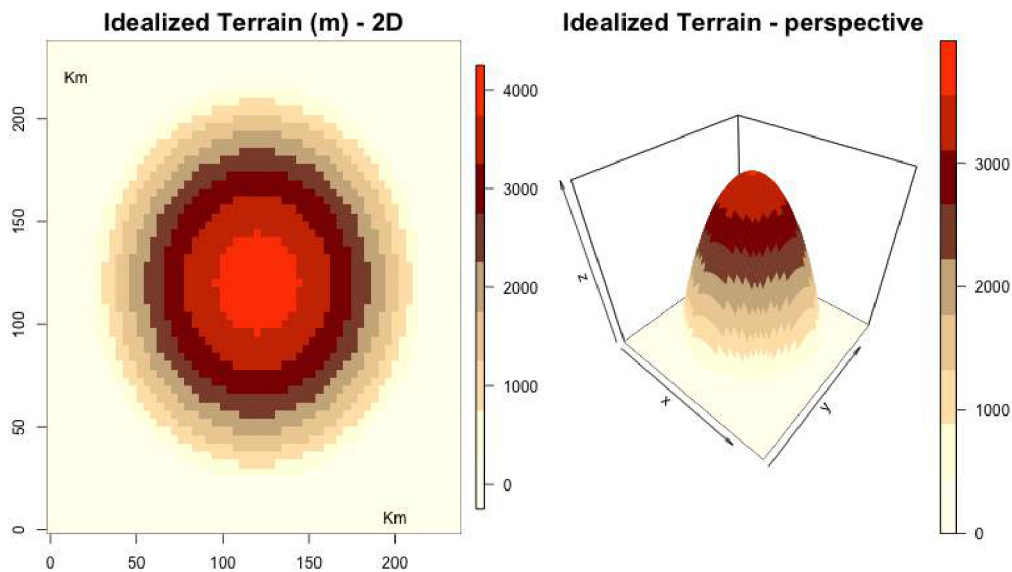
Fig. 3.5: Box plot of number of positions sent per iteration using this scheme

Sub-Experiment - A Fig. 3.6 left panel depict the source function $S(x,y)$ induced by the new terrain and where the quantities are derived for $u = 10m/s$, $v = 10m/s$, $RH = 35\%$

The corresponding simulations form the model are presented in Fig. 3.46

Sub-Experiment - B Fig. 3.6 central panel depicts the source function $S(x,y)$ where the quantities are derived for $u = 15m/s$, $v = 10m/s$, $RH = 35\%$

The corresponding simulations form the model are presented in Fig. 3.47

Sub-Experiment - C Fig. 3.6 right panel depicts the source function $S(x,y)$ where the quantities are derived for $u = 10m/s$, $v = 15m/s$, $RH = 35\%$

The corresponding simulations form the model are presented in Fig. 3.48
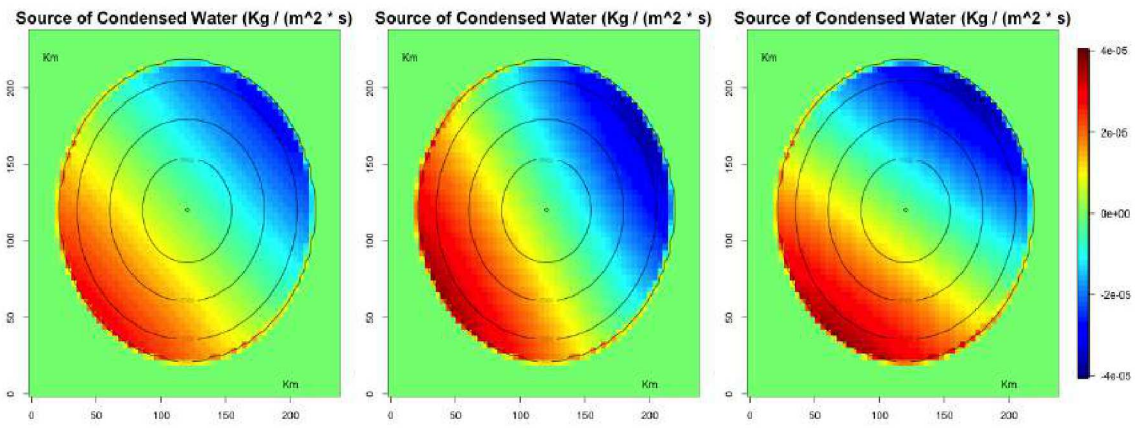
Fig. 3.6: Source function related to the three sub-experiments: A) left panel; B) central panel; C) right panel

Multiple Sequential Ridges

An idealized terrain composed of multiple sequential ridges is built inspired to the work by Colle, 2008. The W-E section is a bell-shaped mountain ridge of height h_m and width a perturbed by a sinusoidal function which creates j peaks before the crest:

East of the crest the terrain elevation is derived from:

$$h(x) = \frac{h_m}{(\frac{x}{a/2})^2 + 1} + h' sin(2\pi j \frac{x}{L}) \tag{3.3}$$

where L is the width of the domain East of the crest. West of the crest the elevation is from:

$$h(x) = \frac{h_m}{(\frac{x}{a/2})^2 + 1} \tag{3.4}$$

$h_m = 2500m$ is the maximum altitude; $a = 50Km$; $j = 2$; $h' = 800m$, is the amplitude of the perturbations.
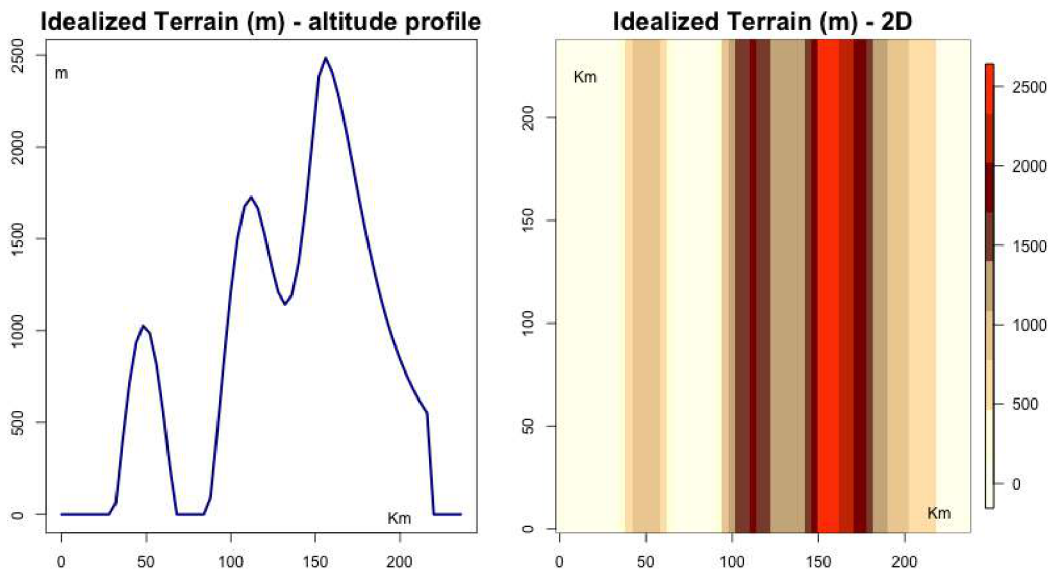
Fig. 3.7: Left: E-W section of the idealized terrain; right: 2D terrain

Sub-Experiment - A Fig. 3.8 left panel depict the source function $S(x, y)$ induced by the new terrain and where the quantities are derived for $u = 10m/s$, $v = 10m/s$, $RH = 35\%$

The corresponding simulations form the model are presented in Fig. 3.49

Sub-Experiment - B Fig. 3.8 central panel depict the source function $S(x, y)$ where the quantities are derived for $u = 15m/s$, $v = 10m/s$, $RH = 35\%$

The corresponding simulations form the model are presented in Fig. 3.50

Sub-Experiment - C Fig. 3.8 right panel depicts the source function $S(x, y)$ where the quantities are derived for $u = 10m/s$, $v = 15m/s$, $RH = 35\%$

The corresponding simulations form the model are presented in Fig. 3.51
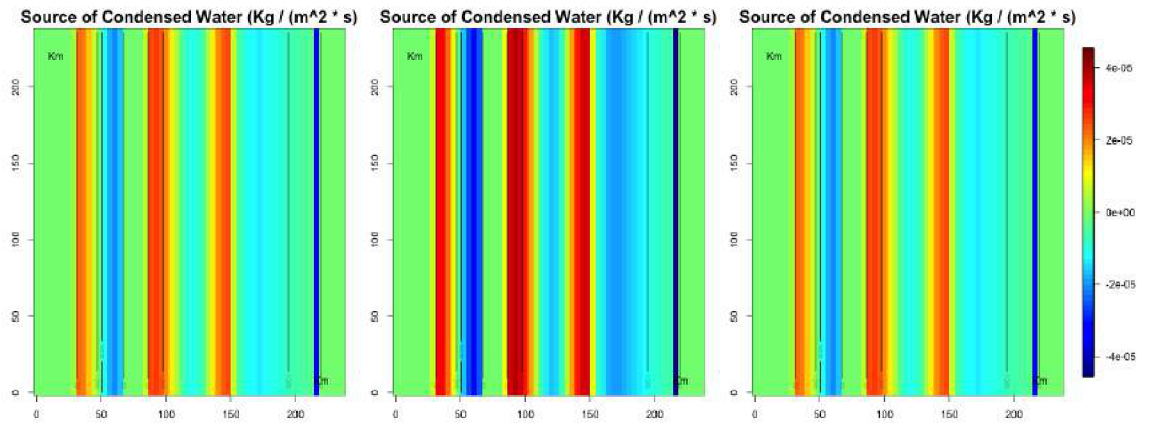
Fig. 3.8: Box plot of number of positions sent per iteration using this scheme

In addition to the comments on the previous experiments, in the presence of multiple ridges, the source of condensation and consequently precipitation is induced on the windward side of each ridge. Each ridge is considered as an independent one, so that the source is built without considering the discharge of rainfall occurred on the preceding slope and is then likely overestimated.

3.2 Case studies on the Coast Range and Sierra Nevada in California, US

3.2.1 Domain and Data

The following case studies are based on observed data of precipitation from CPC US Unified Precipitation data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at http://www.esrl.noaa.gov/psd/. The dataset is composed of daily observations of accumulated precipitation on a $.25 \times .25$ regular grid.

A spatial box centered on California,US is used. The box has corners $-124E, -116E, 35N, 44N$.

The atmospheric variables used are taken from the ERA-Interim reanalysis (Dee et al., 2011). For all the atmospheric fields originally produced on vertical levels in the reanalysis, only the mean value within the vertical interval $850 - 700hPa$ is considered. This vertical layer approximately corresponds to the height of the mountain ranges, and this choice follows the results of Neiman et al., 2002 as already explained in the idealized experiments. The fields at the surface used to build the quantity required in the statistical model are 2 meter temperature and large-scale precipitation. The fields distributed on vertical levels are the horizontal components of the wind, temperature, geopotential and specific humidity.

The terrain elevation is from the GTOPO30 global digital elevation model (DEM) with a horizontal grid spacing of 30 arc seconds from the U.S. Geological Survey's EROS Data Center. In order to enhance the signal and reduce the noise, the DEM was smoothed using a Singular Spectrum Analysis (Ghil et al., 2002), through which the 85% of the variance of terrain altitude is explained.

Fig. 3.9 shows the original terrain and the smoothed one.

3.2.2 Case Study 1: Empirical Estimation

The first case study refers to the 24 hours spanning from 12:00 UTC Dec,1 2005 to 12:00 UTC Dec,2 2005. The atmospheric conditions recorded in the selected 24h interval created a typical interaction between synoptic westerly frontal systems moving to the continent and the orography of the Coast Range and Sierra Nevada.

Figure 3.10 shows the observed precipitation on the $.25 \times .25$ degrees lattice composed of $33 \times 33 = 1089$ points. The right panel shows a histogram for the intensity of recorded precipitation.

A site is usually considered "wet" if the recorded precipitation exceeds $1mm$. The day considered in this case study can be seen as a prevalently "wet" day: among the
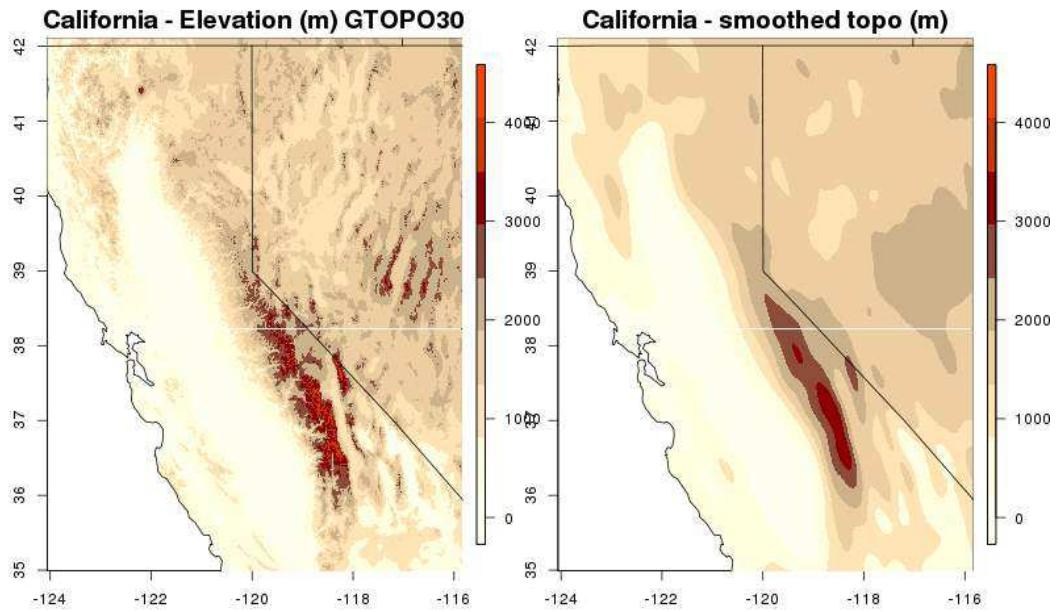
Fig. 3.9: Spatial domain - Left: Elevation GTOPO30; Right: smoothed terrain.

959 sites on land, 815 recorded at least $1mm$ of rainfall during the 24 hours, for a fraction of 84% of "wet" sites. This disproportion might create a situation of *model overfitting*: let us imagine we want to improve the agreement of the model with the observations in terms of occurrence. Such a disproportion will likely produce an increasing in agreement simply by increasing the number of sites predicted as "wet", regardless the position and a balance between the true classification of the mistaken sites. To enhance the signal, a threshold of $2mm$ was instead chosen to divide "dry" from "wet" sites, which creates a percentage of 78% of "wet" sites on the area.
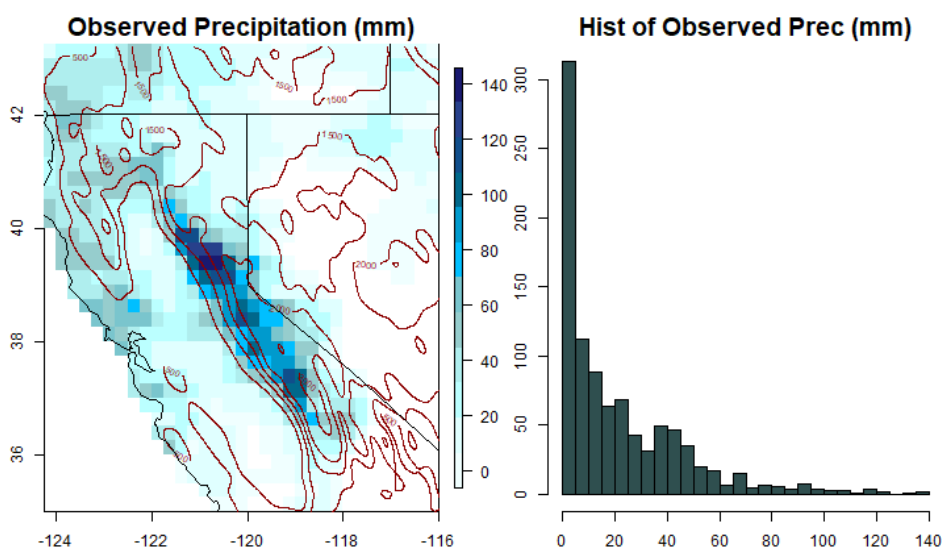
Fig. 3.10: Observed Precipitation: CPC US Unified Precipitation data

The observed precipitation is assumed to be the result of both large scale processes and orographic effects. The fraction due to the large-scale precipitation is taken from the ERA-Interim reanalysis, as shown in Fig. 3.11.
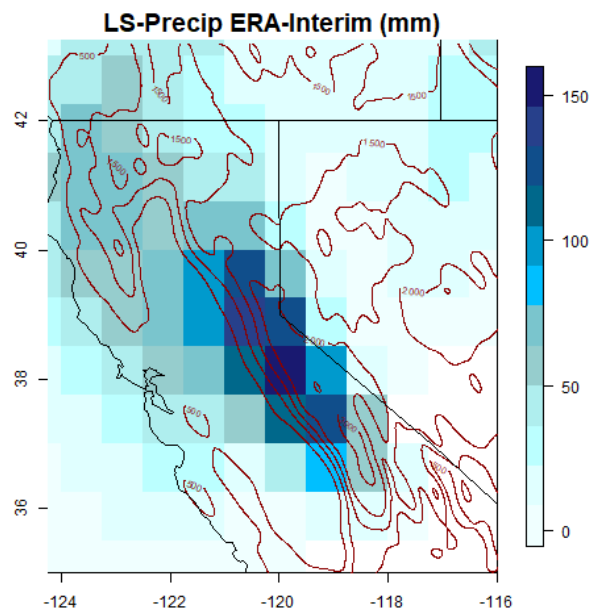
Fig. 3.11: Large Scale precipitation fraction - ERAInterim

The static stability was estimated following Smith and Barstad, 2004 from ERA-Interim data by the ratio $\frac{\gamma_{mo}}{\gamma}$, where the numerator is the moisture lapse rate and the denominator is the environmental lapse rate. Values greater than 1 indicate a static stable atmosphere. The saturation water vapor density ρ_{sref} was computed using the *Tetens equation* for the saturation vapor pressure. The saturation water vapor pressure and the static stability are shown in Fig. 3.12 and 3.13 respectively. In both figures the wind field is overimposed by means of arrows. Average wind components advecting cloud water are $u = 9.752 m/s$ and $v = 7.005 m/s$
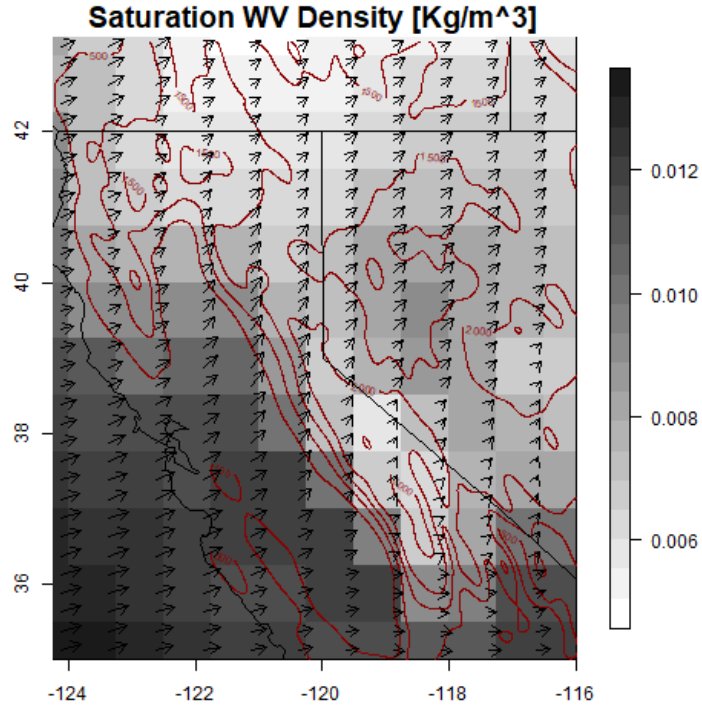
Fig. 3.12: Saturation water vapor density $[Kg/m^3]$; Arrows: horizontal wind

Results

The model in 2.33 with the inclusion of the large-scale precipitation is the one applied in this case study:

$$p^* = \alpha * LSP + Wp^* + S + \eta \tag{3.5}$$

with mean $\mu = (I - \tau \tilde{W})^{-1}(\alpha LSP + S)$ and covariance matrix $\Sigma = \sigma^2 (I - \tau \tilde{W})(I - \tau \tilde{W}^T)$. The variance of the noise term η was set as $\sigma^2 = 1e - 10$. The value of the
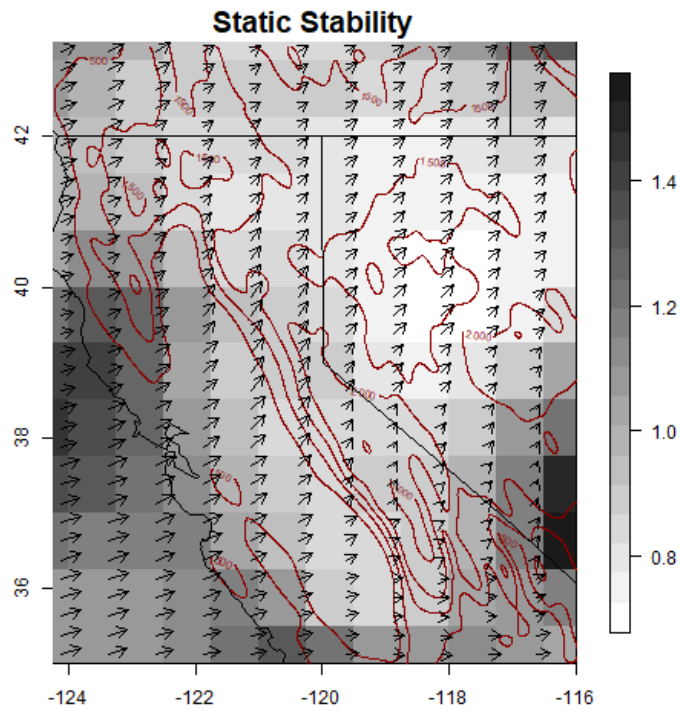
Fig. 3.13: Estimate of static stability as γ_m/γ. Values greater than 1 refers to stable atmosphere. Arrows: horizontal wind

latter was set because for model identifiability it is not possible to estimate both σ^2 and α at the same time .

Fig. 3.14 shows the source function of condensation $S(x,y)$ computed according to eq. 2.3 :
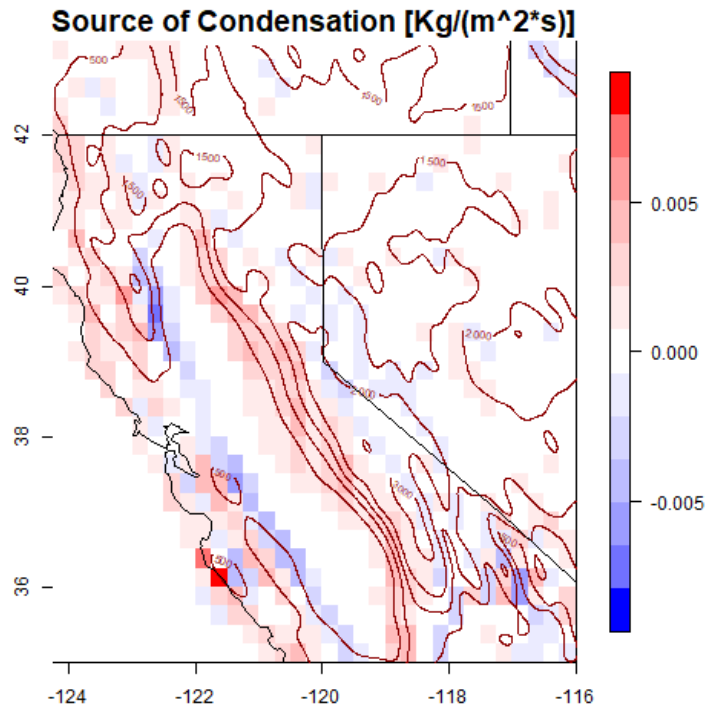
Fig. 3.14: Source of condensation arising from forced ascent of the moist flow on the mountain range, $\left[\frac{Kg}{m^2*s}\right]$

The present case study can serve to show in an empirical way the necessity of including the additional term LSP representing large-scale precipitation. This requirement is evident comparing the observed spatial pattern of precipitation and the pattern produced by the model ignoring the large-scale precipitation (setting $\alpha = 0$). The results of this test are shown in Fig. 3.15

In particular, the simulation for the test was performed letting τ vary between $400s$ and $4000s$. In all simulations we can observe the orographic effect on the precipitation pattern at the ground, namely an increase of precipitation on the windward side of the range and dryer conditions in the lee sides. Completely dry lee sides are in disagreement with the observations, according to which precipitation also occurs East to the mountain ranges (both Coast Range and Sierra Nevada), although less intense. That is true because of other mesoscale processes not accounted in the model. These processes can be taken into consideration by including the Large-scale component of the precipitation from ERA-Interim data as in the formulation in eq. 2.18. The orographic processes act then as alteration and reinforcement of the large-scale processes.

Given the importance of including large-scale precipitation, in the case study here presented the value of the parameters τ and α were estimated as first approach using
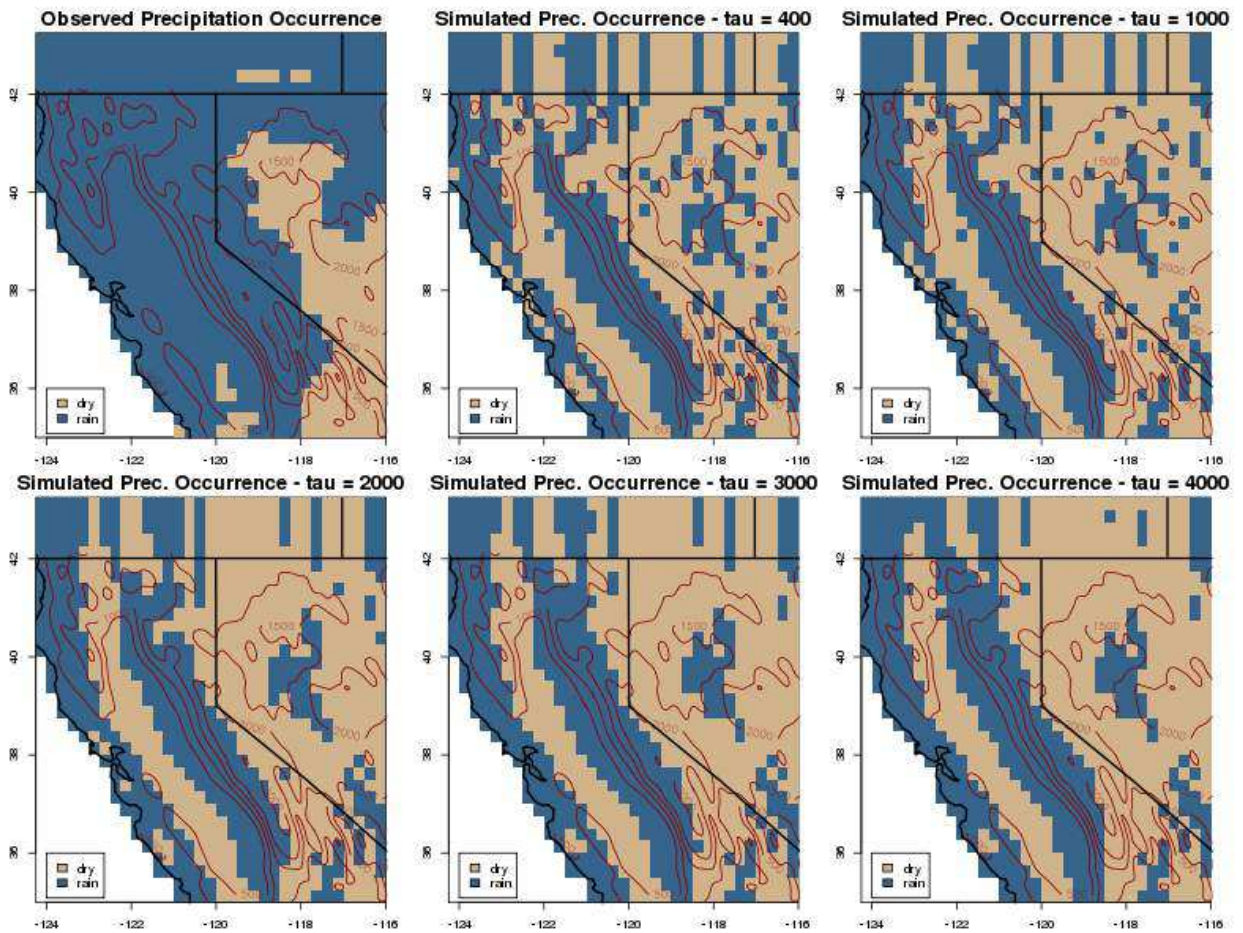
Fig. 3.15: Simulations of precipitation occurrence letting τ vary.

an empirical method for maximizing the accuracy of the estimates of precipitation occurrence over the 1089 sites. In particular, a grid of possible values for both τ and α was chosen and the results of the simulation in terms of occurrence, obtained prescribing the couples of values for the parameters, were compared with the observations. The couple of values which guarantees the best accuracy was considered as the best estimate of the unknown parameters.

Precipitation Occurrence

For the occurrence the best set of parameters found were $\hat{\alpha} = 11, \hat{\tau} = 3600s$. The mean term $\mu = (I - \tau\tilde{W})^{-1}(\alpha LSP + S)$ contains a sum of contributions of a proportion of LSP given by α and S. Thus, the estimated value $\hat{\alpha} = 11$ means the contribution of large scale precipitation dominates the determination of the spatial distribution of wet sites. Fig 3.16 shows the simulated latent gaussian *precipitation potential* obtained prescribing $\hat{\alpha} = 11, \hat{\tau} = 3600s$, and Table 3.2.2 summarizes the results, showing an overall accuracy of 84% Fig. 3.17 shows the simulated field prescribing $\alpha = 11, \tau = 3600s$.
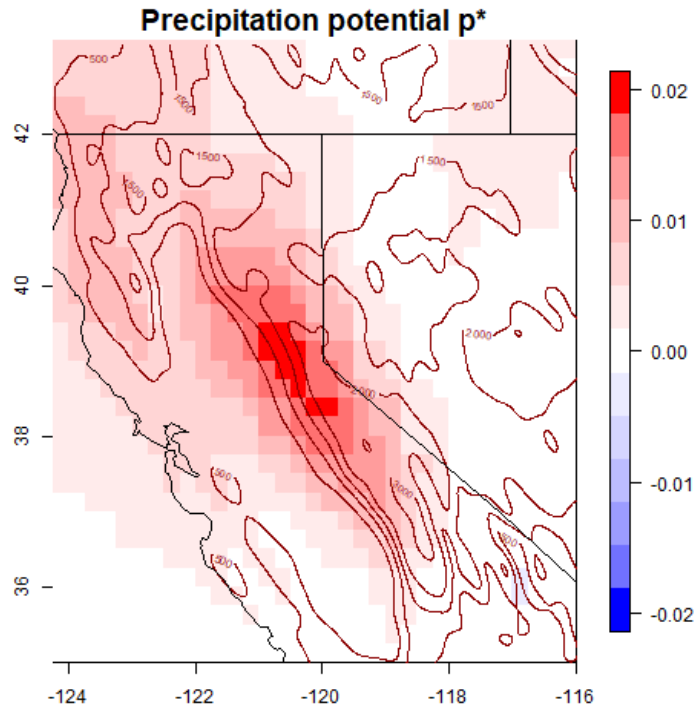
Fig. 3.16: Simulation of Precipitation Potential p* with $\alpha = 11, \tau = 3600s$
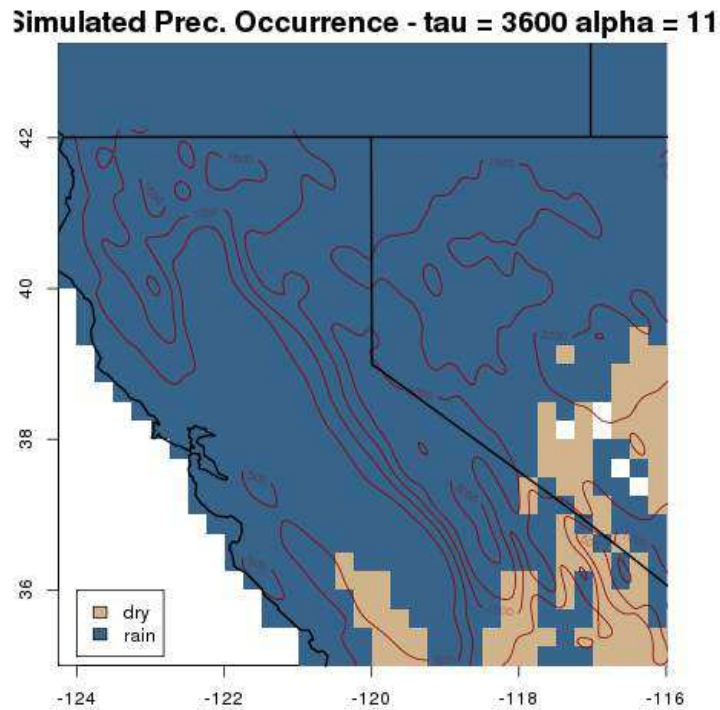
Fig. 3.17: Simulation of Precipitation Occurrence with $\alpha = 11, \tau = 3600s$

	simulated	
observed	no rain	rain
no rain	68	131
rain	22	738

Precipitation Intensity

To determine the intensity of precipitation, an approach based on Katz and Parlange, 1995 was used: the intensity is found applying a power transform of order $k, 0 < k < 1$ to the latent gaussian field p^* on "wet" sites, when $p^* > 0$. k is chosen here as the value that makes the $95th$ percentile of both the distributions be equal. Namely, letting y be the observed precipitation on "wet" sites:

$$k : (p^*)^k_{[.95]} = y_{[.95]} \qquad (3.6)$$

The use of the $95th$ percentile allows to shrink most of the distribution of the positive values of p^* making it match the observed distribution but without relying on the very highest observed values which might be outliers. This approach also implies that the spatial dependence is only explained by the latent field p^* and the power transform is homogeneous over the area.

The latent field p^*, obtained setting $\alpha = 11, \tau = 3600$ as estimated in the previous section, was transformed into precipitation intensity after finding $k = 0.62$. The results are presented in Fig. 3.20.

To evaluate the model skill, two global measures, namely the Root Mean Square Error (RMSE) and Pattern Correlation (PC) are computed:

$$\text{RMSE} = 20.998$$
$$\text{PC} = 0.7955$$

The simulation underestimates the highest observed values, but an overall overestimation of both precipitation occurrence and intensity is evident. In particular, the signal due to the orography and visible in the observed data seems to be hidden by large-scale precipitation. This most likely happens because the accuracy in terms of occurrence was favored and because of the disproportion between wet and dry sites (see beginning of Chapter 3.2). In this situation, a different quantity might be minimized for the choice of α, τ: the quantity was chosen as the weighted sum of the errors in the classification (number of false dry, number of false wet) with weights inversely proportional to the number of observed dry or wet sites. In particular,
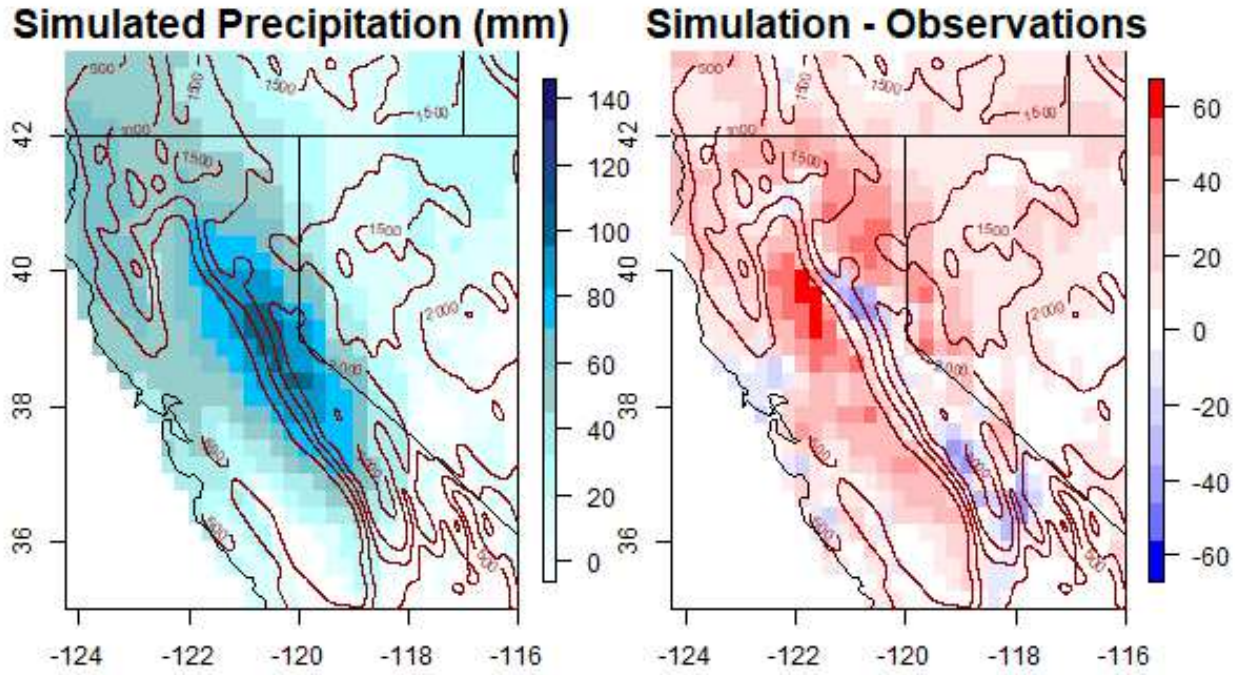
Fig. 3.18: Simulated precipitation intensity with $\alpha = 11, \tau = 3600$ (left) and observations (right)

letting w_d be the number of sites observed as dry and classified as wet, and d_w the opposite, the quantity to be minimized in the second experiment is:

$$C = w_d * (1 - \frac{n.dry}{n}) + d_w * (1 - \frac{n.wet}{n}) \tag{3.7}$$

where $n.wet$ and $n.dry$ are the number of wet and dry sites respectively and n the total number of sites. This choice forces the model to represent well the less represented class. It is expected to balance the representation of both the large scale precipitation and the orographic effects.

The values $\hat{\tau} = 1600, \hat{\alpha} = 3$ minimize the quantity C. Fig. 3.19 shows the latent *precipitation potential* simulated with those values of parameters, and table 3.2.2 reports the accuracy of the results in terms of occurrence, with an overall accuracy of 81.1%. The value k was recomputed, and the best estimate identified as $k = 0.71$. Fig 3.20 shows the simulation of precipitation intensities obtained from the new set of values for the parameters and its comparison with observations. Global measures of skill computed for this experiment are again RMSE and Pattern Correlation:

$$\text{RMSE} = 16.12$$
$$\text{PC} = 0.783$$
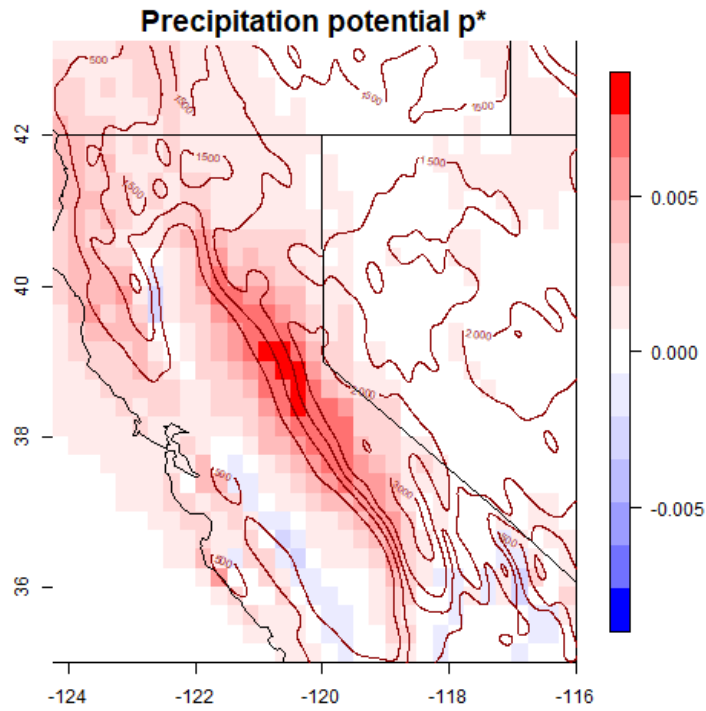
Fig. 3.19: Simulation of precipitation potential p* for $\hat{\tau} = 1600, \hat{\alpha} = 3$
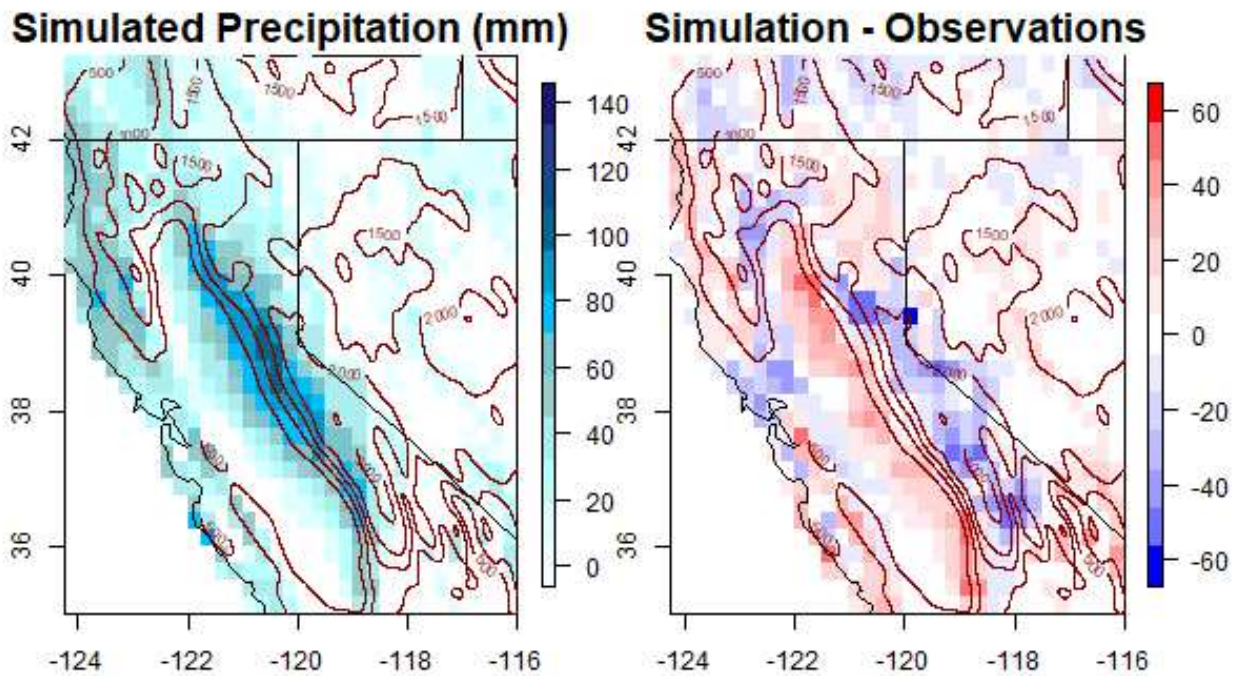
Fig. 3.20: Simulation of precipitation intensity for $\hat{\tau} = 1600, \hat{\alpha} = 3, k = 0.71$ and comparison with observations

	simulated	
observed	no rain	rain
no rain	123	76
rain	187	573

The new criterium C chosen to determine the best values of α and τ improves the simulation: the spatial pattern are better represented and both the orographic precipitation and the large-scale precipitation are balanced and show good agreement with the observations. Moreover the underestimation of the highest observed values is less present.

3.2.3 Case Study 2: MCMC estimation

Controlled experiment

A controlled experiment is carried out in order to understand the properties of the MCMC method presented in section 2.3.1 in a quasi-real setting. The atmospheric conditions are those recorded in the 24hr interval from 12:00 UTC Feb,5 2006 to 12:00 UTC Feb,6 2006 over a spatial box centered over Sierra Nevada with corners -122E, -116E, 35N, 40.5N. The domain is smaller than the one considered in the previous experiment. The terrain elevation and the smoothed topography are shown in Fig. 3.21 in the left and right panel respectively. Furthermore, a gray box indicates the study area in the right panel.
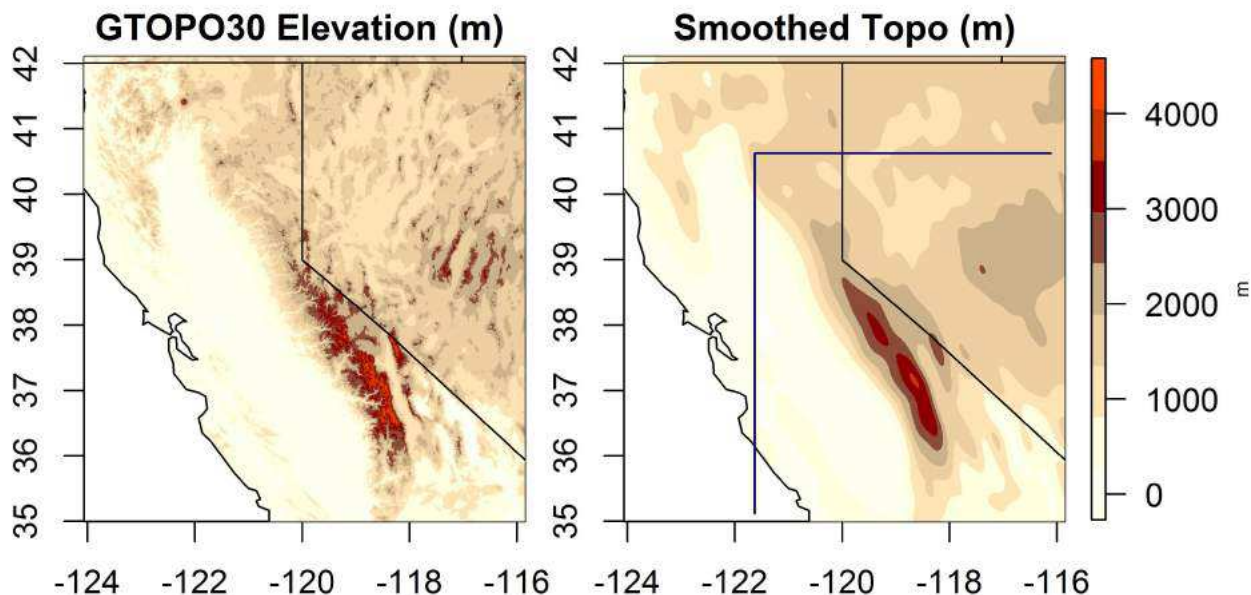
Fig. 3.21: Left Panel: GTOPO30 elevation over California (m); Right panel: smoothed topography (m) by SSA retaining the 85 of variance and a gray square showing the actual area where the second casy study is carried out.

The model version of equation 2.34 is used in this experiment, prescribing $\tau = 5000, \lambda_1 = 0.6, \lambda_2 = 4$ for the three unknown parameters. Fig. 3.22 shows the Source of condensation and Large scale precipitation from ERA-interim, while Fig. 3.23 shows the latent precipitation potential $p*$ and the occurrence simulated.
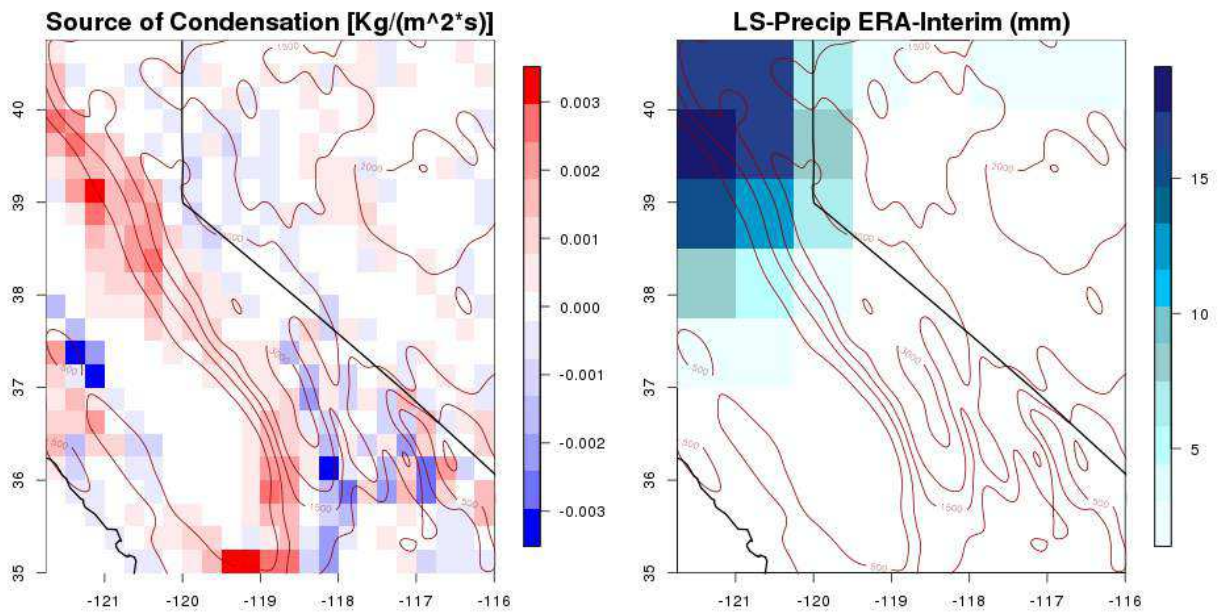
Fig. 3.22: Left Panel: Source of cloud production; right panel:

In order to understand the behavior of the Markov Chains, 20 parallel chains ran from overdispersed initial values for a very high number of iterations, namely 100000. Initial values range is $\tau : [400, 6000], \lambda_1 : [0.1, 5], \lambda_2 : [0.1, 5]$. Fig. 3.24 shows the results:

The next table summarizes the results showing for each paramater the true value, the estimated value from the posterior mean and the posterior standard deviation.
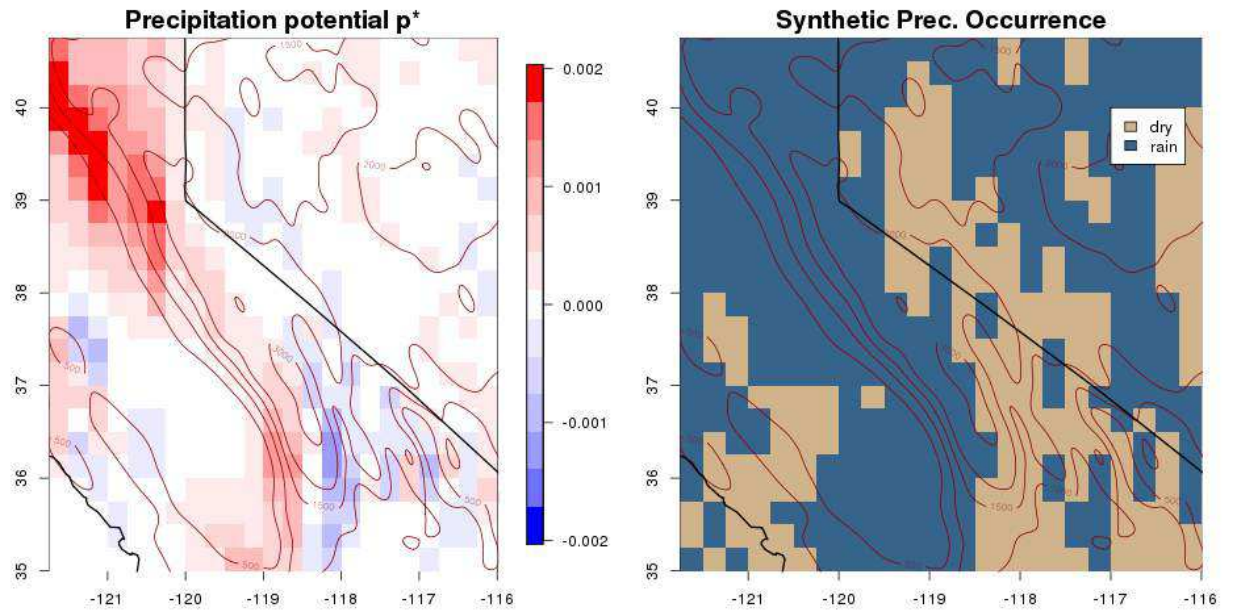
Fig. 3.23: Left Panel: precipitation potential simulated from $\tau = 5000$, $\lambda_1 = 0.6$, $\lambda_2 = 4.$; right panel: simulated occurrence.

Parameters	True Parameter Value	Posterior Means	Posterior Standard Deviation
τ	5000	4605.69	1206
λ_1	0.6	0.63	0.38
λ_2	4	4.63	2.75
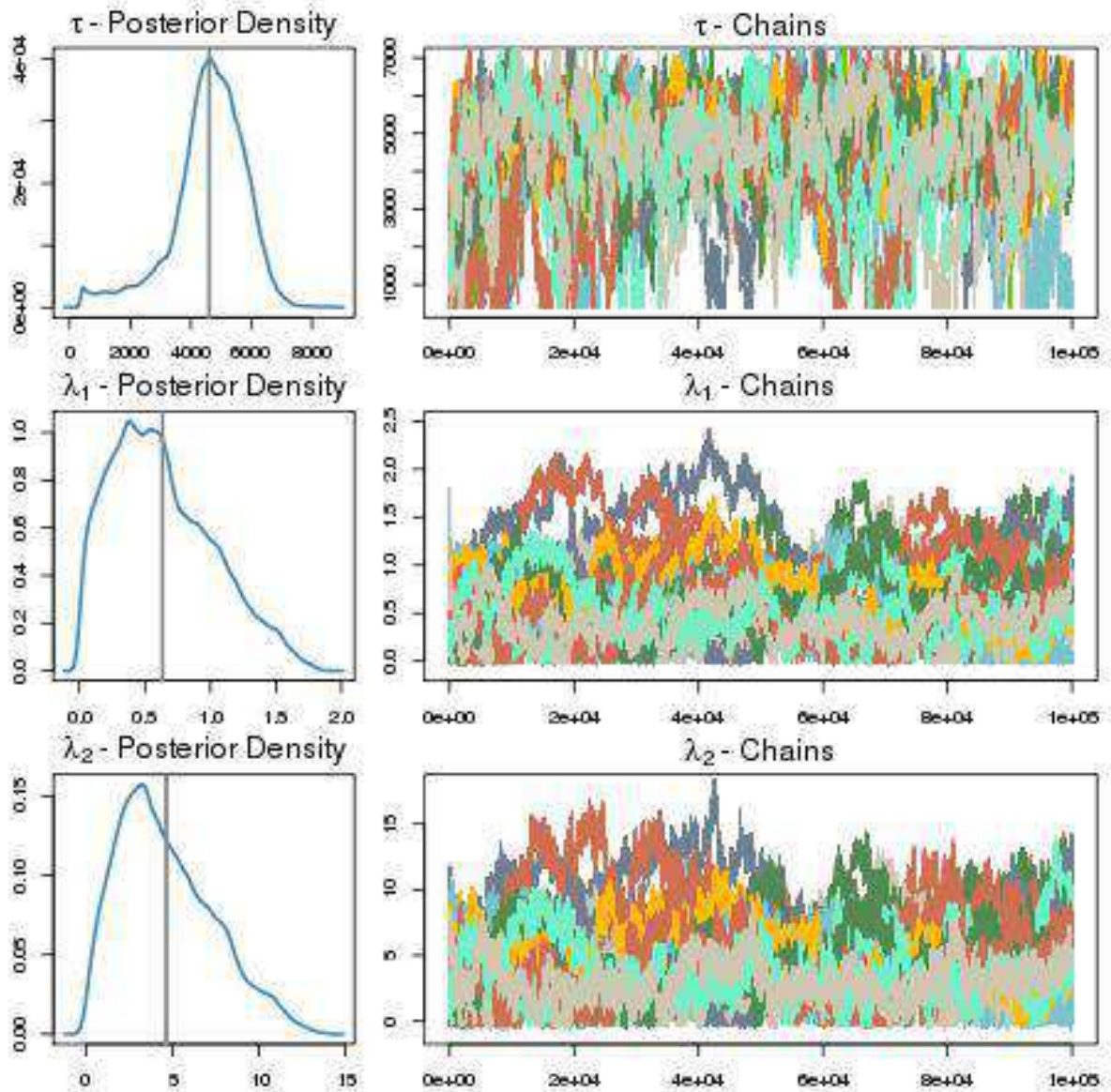
Fig. 3.24: Controlled experiment - First columns of panels show the posterior densities of tau (first row), λ_1 (second row), λ_2 (third row). The densities are estimated using only iterations from 50000 to 100000. The gray lines indicate the posterior means. Plots in the right column: 20 chains for the three parameters.

Case Study on real data - A: 12:00 UTC Dec,1 2005 to 12:00 UTC Dec,2 2005

The second case study is performed on the time interval from 12:00 UTC Dec,1 2005 to 12:00 UTC Dec,2 2005. That is the same day considered in the case study in section 3.2.2 in which the parameters were estimated empirically. The spatial box is still centered over the Sierra Nevada, but is smaller, with corners -122E, -116E, 35N, 40.5N. Terrain elevation and the smoothed topography are shown in Fig. 3.21 in the left and right panel respectively. Observations and atmospheric conditions are depicted in Fig. from 3.10 to 3.13. Model 2.34 is used in this case study.

Parameter estimates are found following the method based on MCMC outlined on section 2.3.1. The results from 7 parallel chains, initialized with different start values are shown in Fig 3.25

The next table summarizes the results showing for each parameter the estimated value from the posterior mean and the posterior standard deviation.

Parameters	Posterior Means	Posterior Standard Deviation
τ	4551	2525
λ_1	0.00038	0.0004
λ_2	0.028	0.0068

Fig. 3.26 shows the precipitation potential field simulated from the parameters value estimates presented in the previous table:

Accuracy on precipitation occurrence is displayed in the following confusion matrix:

	Simulated 0	Simulated 1
Observed 0	74	73
Observed 1	37	345

The following Fig. 3.27 presents accumulated precipitation obtain from precipitation potential applying the power transformation of eq. 2.27.

Global skill with respect to observations was assessed by means of the RMSE and Pattern Correlation:

$$\text{RMSE} = 17.005$$
$$\text{PC} = 0.84$$
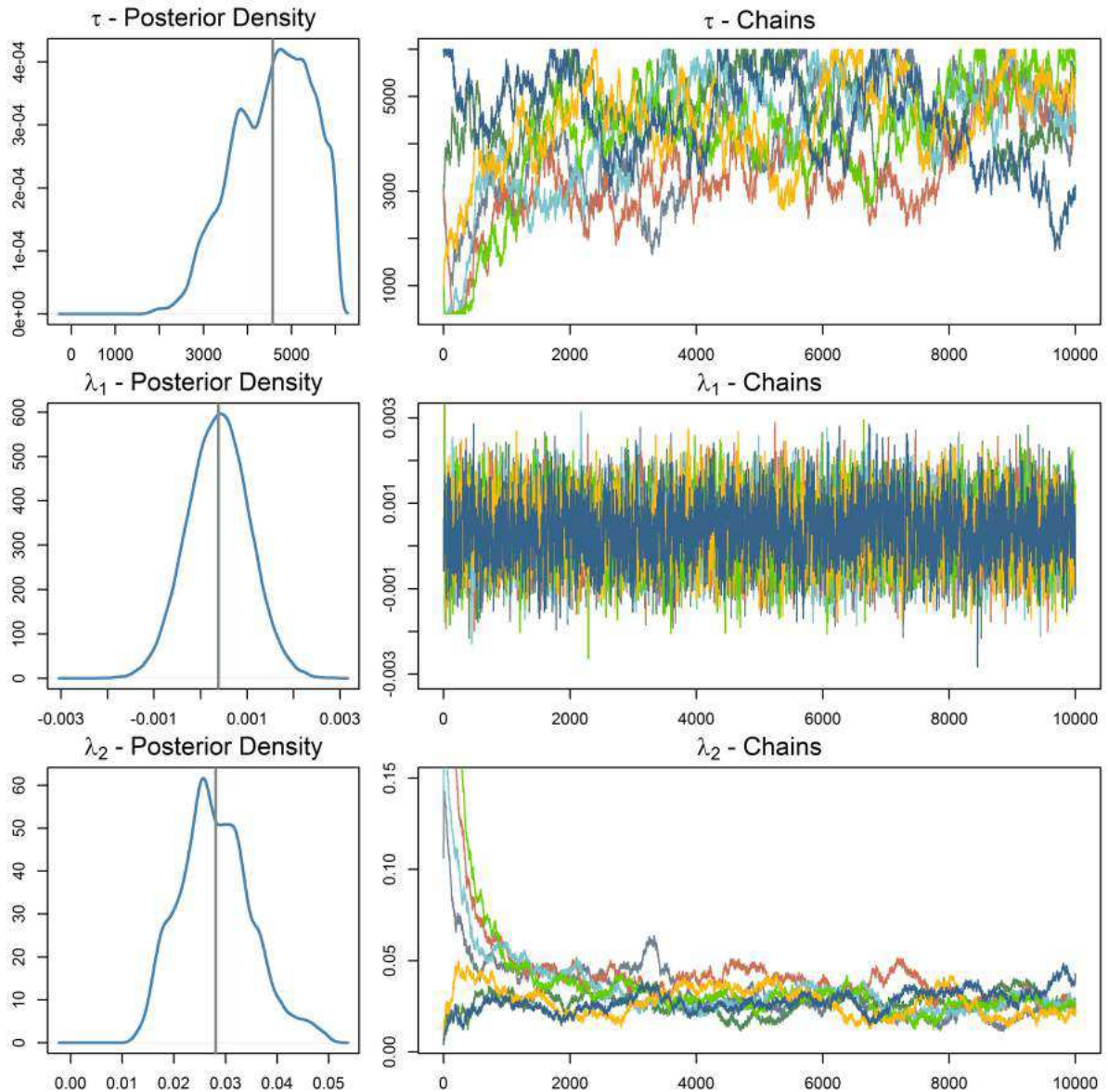
Fig. 3.25: Experiment on the 2nd Dec: 7 Monte Carlo Chains initialized with different values. Panels in the first column show the posterior densities for τ, λ_1 and λ_2. Panels in the second column show trace plots of the chains for the same three parameters.

Case Study on real data - B: 12:00 UTC Feb,5 2006 to 12:00 UTC Feb,6 2006

The third case study is performed on the 24hr interval from 12:00 UTC Feb,5 2006 to 12:00 UTC Feb,6 2006. The spatial box is the same depicted in Fig. 3.21 with corners -122E, -116E, 35N, 40.5N. Fig. 3.28 presents the observed 24hr accumulated precipitation on the area . With respect to Case Study A, precipitation intensity is lower as well the area over which precipitation occurs is smaller.
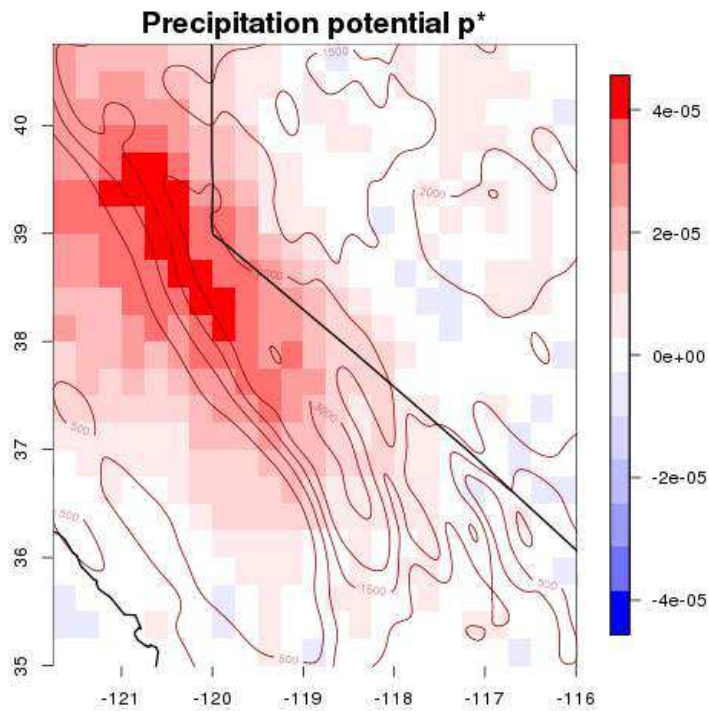
Fig. 3.26: Experiment on the 2nd Dec: simulated precipitation potential from parameter estimates $\tau = 4551, \lambda_1 = 0.00038, \lambda_2 = 0.028$.
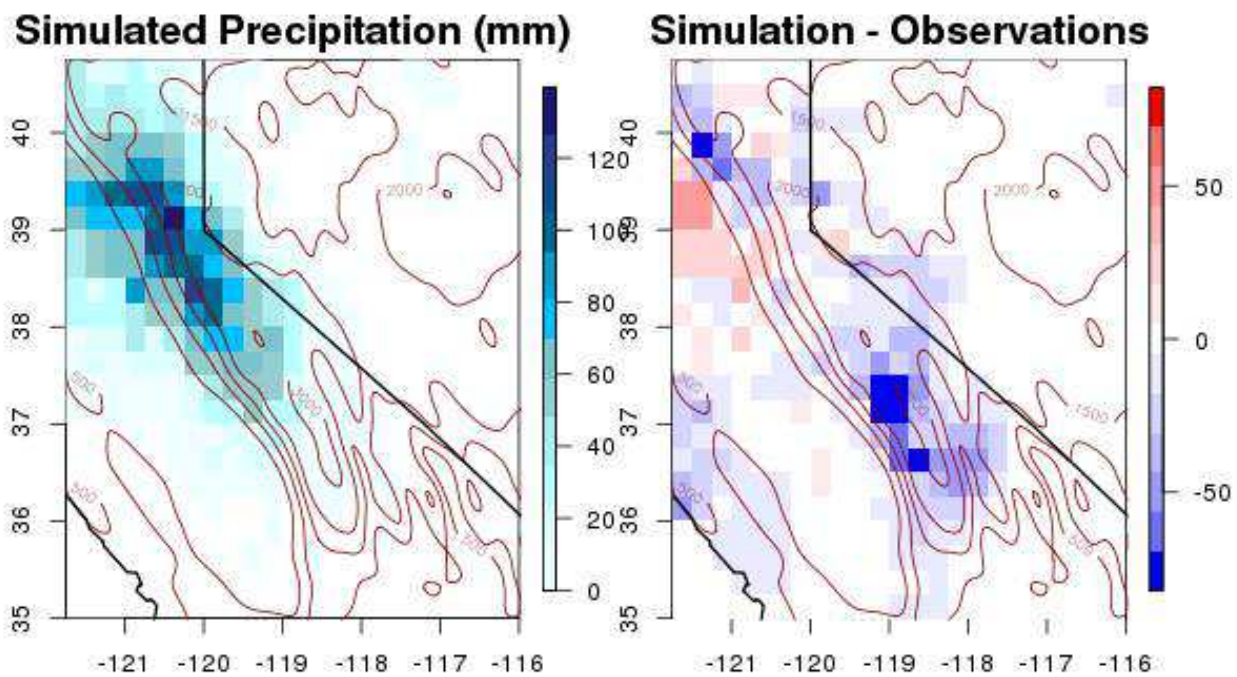
Fig. 3.27: Experiment on the 2nd Dec: simulated precipitation intensity and comparison with observations
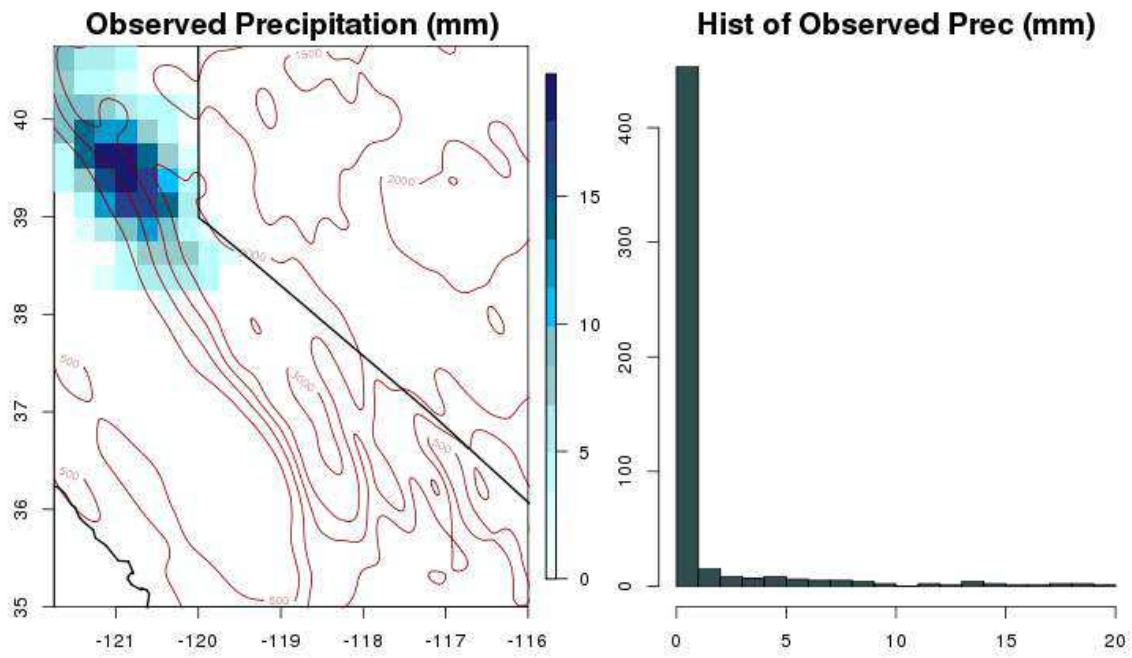
Fig. 3.28: Left Panel: 12:00 UTC Feb,5 2006 to 12:00 UTC Feb,6 2006 accumulated precipitation; Right panel: histogram of values of accumulated precipitation over the area

Fig. 3.29 shows the source S of condensation due to local orographic processes and Fig. 3.22 right panel is the large-scale precipitation component from ERA-interim on the time interval considered.

3 chains were used to estimate the parameters. Their traceplots and the constructed posterior densities are shown in Fig. 3.30

The next table summarizes the results showing for each parameter the estimated value from the posterior mean and the posterior standard deviation. Fig. 3.31 presents the precipitation potential p* simulated from those parameter estimates.

Parameters	Posterior Means	Posterior Standard Deviation
τ	5923.104	108.69
λ_1	0.001	0.00071
λ_2	0.0195	0.00867

Accuracy on precipitation occurrence is displayed in the following confusion matrix:

	Simulated 0	Simulated 1
Observed 0	381	72
Observed 1	10	64
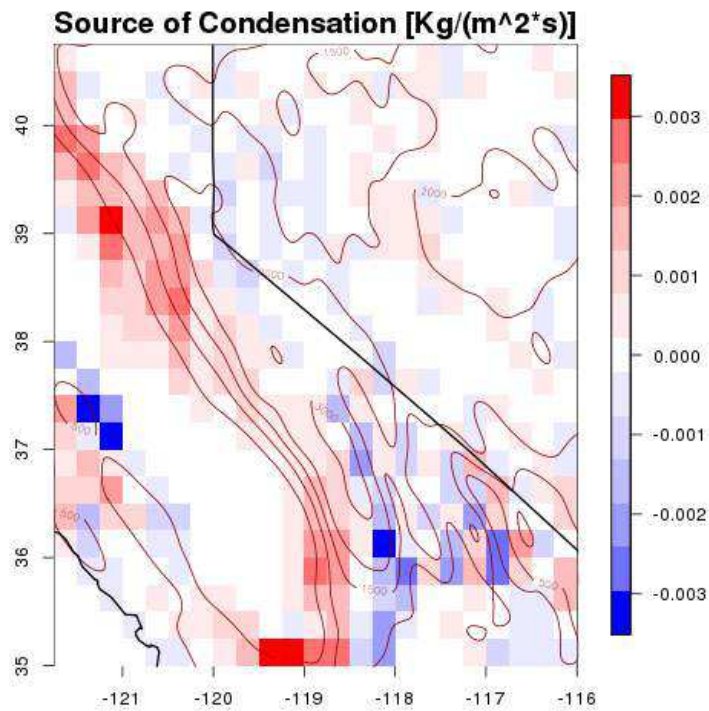
Fig. 3.29: Source of cloud water: 12:00 UTC Feb,5 2006 to 12:00 UTC Feb,6 2006;

The simulated precipitation intensity is shown in Fig. 3.32 together with a comparison with observations. Global measure of skill computed for precipitation intensities are RMSE and Pattern Correlation:

$$RMSE = 1.63$$
$$PC = 0.845$$
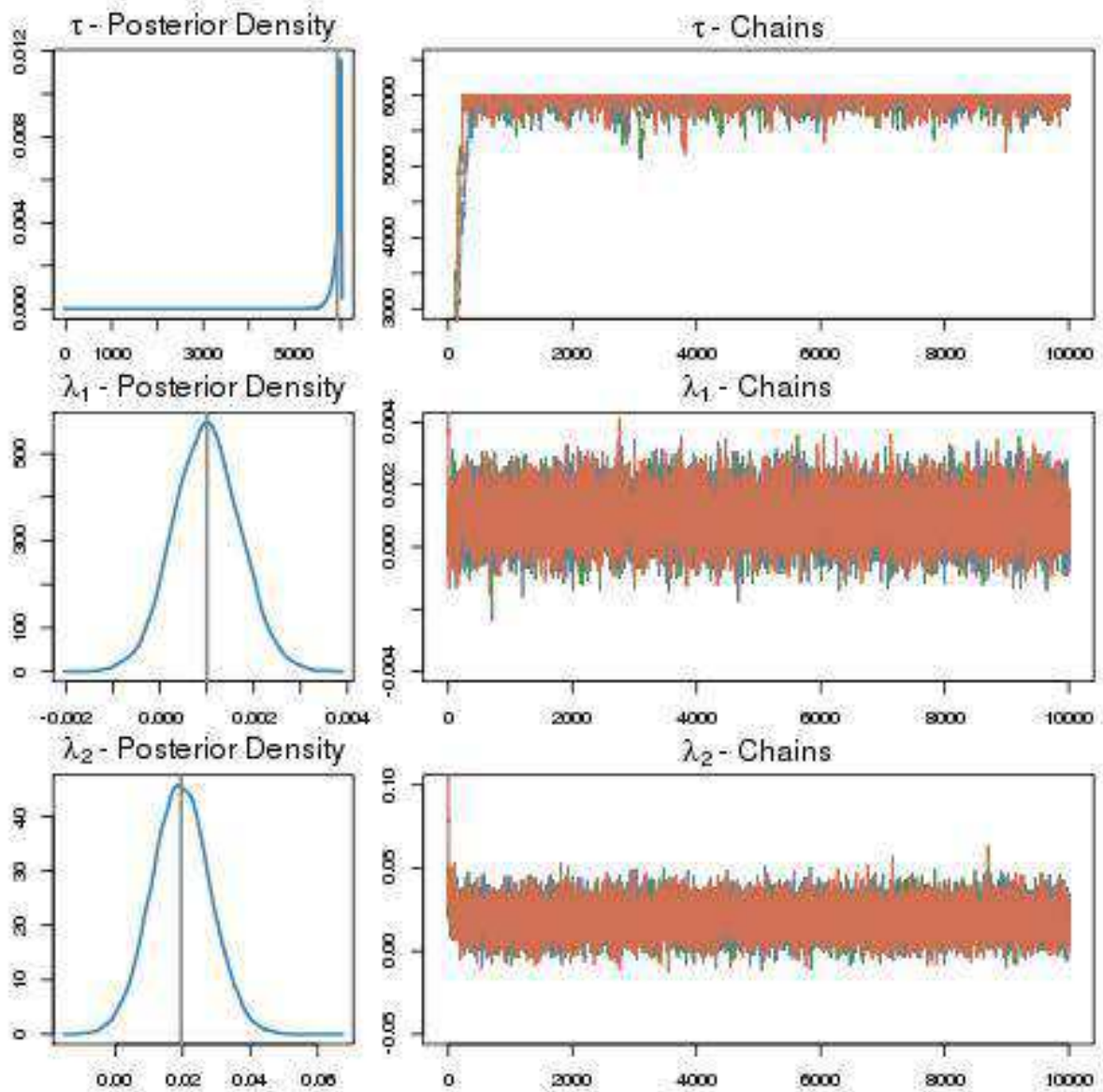
Fig. 3.30: Experiment on the 5th February 2006: 3 Monte Carlo Chains initialized with different values. Panels in the first column show the posterior densities for τ, λ_1 and λ_2. Panels in the second column show trace plots of the chains for the same three parameters.

3.2.4 Case Study 3: Unevenly spaced data

This section presents a case study carried out with unevenly spaced observations on the time interval from 12:00 UTC Dec,1 2005 to 12:00 UTC Dec,2 2005 , the same time span considered in experiment A presented the previous section.

Observations are station-data daily accumulated precipitation from the Global Historical Climatology Network (GHCN) dataset, provided by the National Oceanic and At-
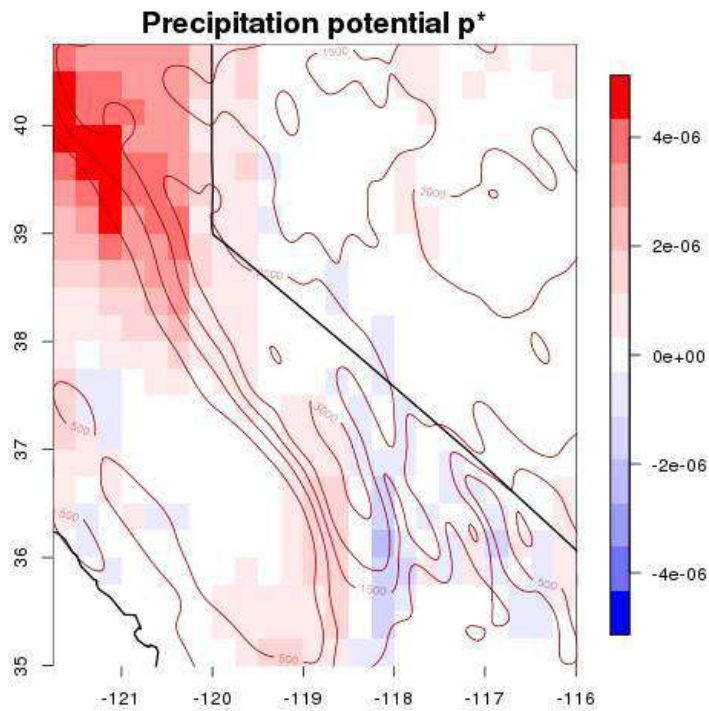
Fig. 3.31: Experiment on the 5th February 2006: precipitation potential simulated from $\tau = 5923, \lambda_1 = 0.001, \lambda_2 = 0.0195$
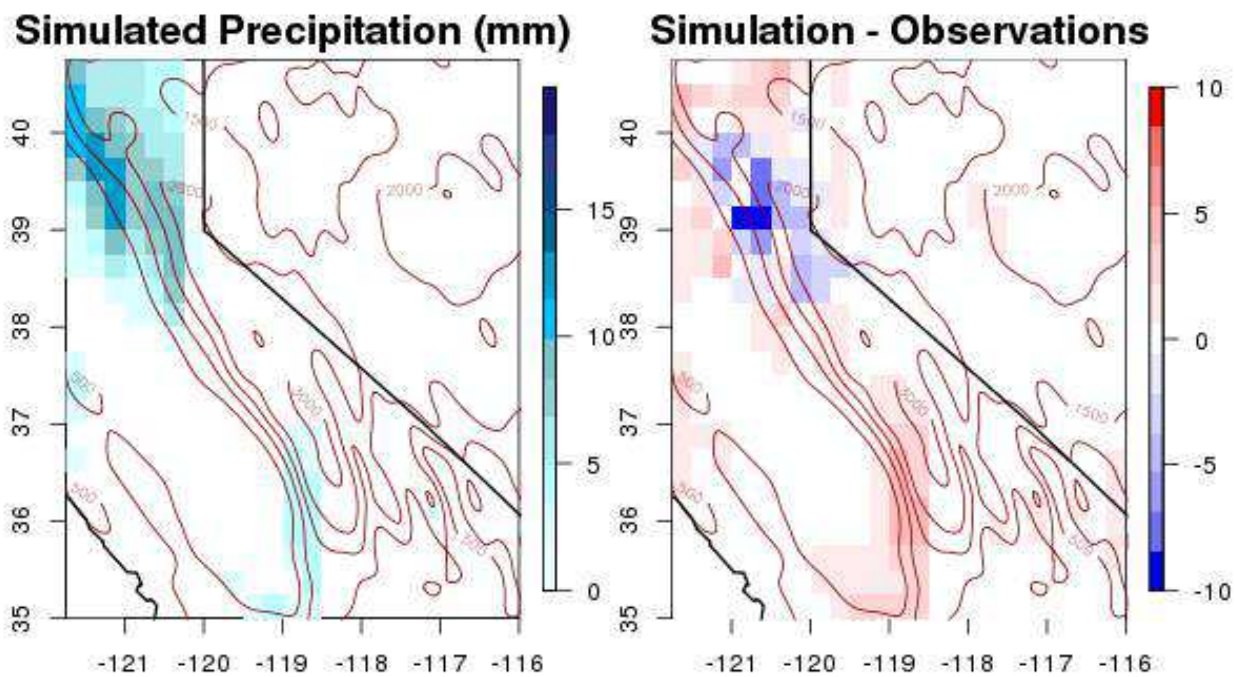
Fig. 3.32: Case Study B, 5 February 2006: Simulated precipitation amount and comparison with observations

mospheric Administration (NOAA) and retrievable from their website (https://www.ncdc.noaa.gov/ghcn-daily-description).

On the day of interest, precipitation data from 207 stations were available over the study region, as shown in Fig. 3.33.
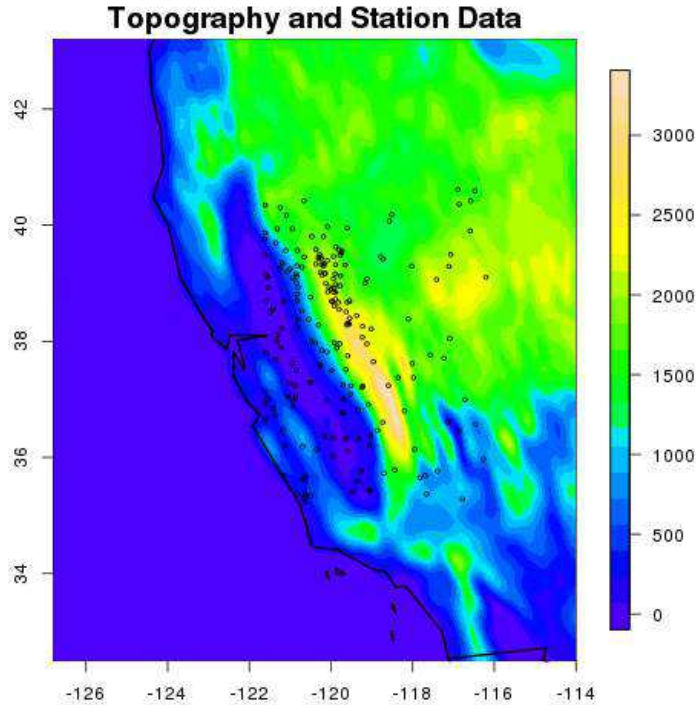
Fig. 3.33: Case Study on unevenly spaced data: 2 December 2005: Topography and location of station data.

Matrix W was built as explained in section 2.2.4. 3 chains were used to estimate the unknown parameters. While convergence appear clear for parameter λ, convergence of τ is more questionable. Fig. 3.34 reports the posterior densities of the parameters and the trace plots, and the table summarizes the posterior means and standard deviations.

Parameters	Posterior Means	Posterior Standard Deviation
τ	417.66	37.33
λ_1	0.00065	0.00051
λ_2	0.005	0.001

Accuracy on precipitation occurrence over the 207 observations is displayed in the following confusion matrix:

	Simulated 0	Simulated 1
Observed 0	14	41
Observed 1	9	142
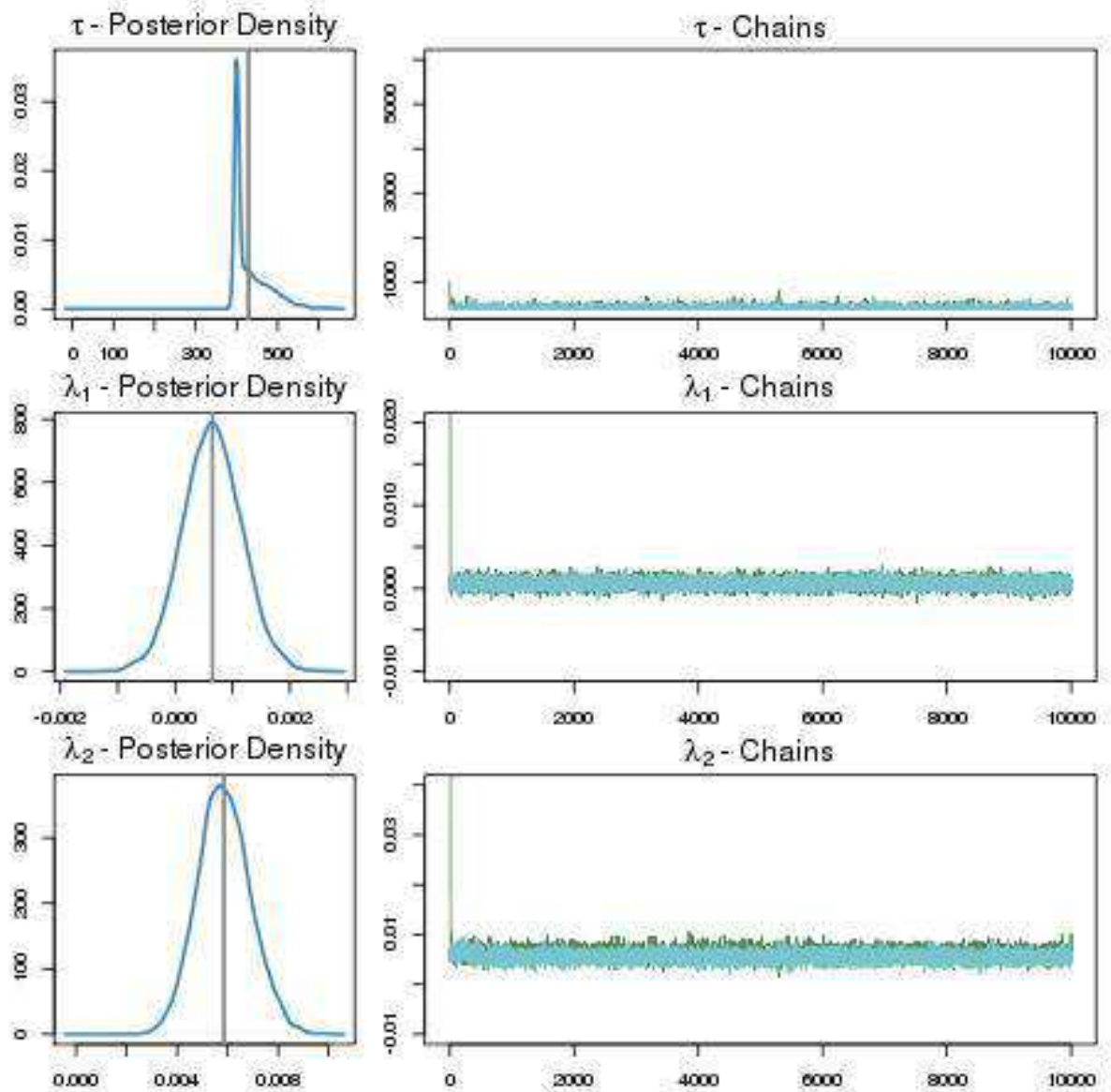
Fig. 3.34: Case Study on unevenly spaced data: 2 December 2005: Trace plot of Markov Chains and posterior densities.

After the power transform was applied to the latent field p^* to obtain the intensities, the following measures of skill were computed:

$$\text{RMSE} = 33.753$$
$$\text{PC} = 0.5$$

Fig. 3.35 presents the spatial latent field p^* obtained on the regular grid used in the previous case studies A, B after the parameters values are estimated from the unevenly spaced data. Fig. 3.36 reports the precipitation intensities obtained from the latent p^* after applying the power transform.
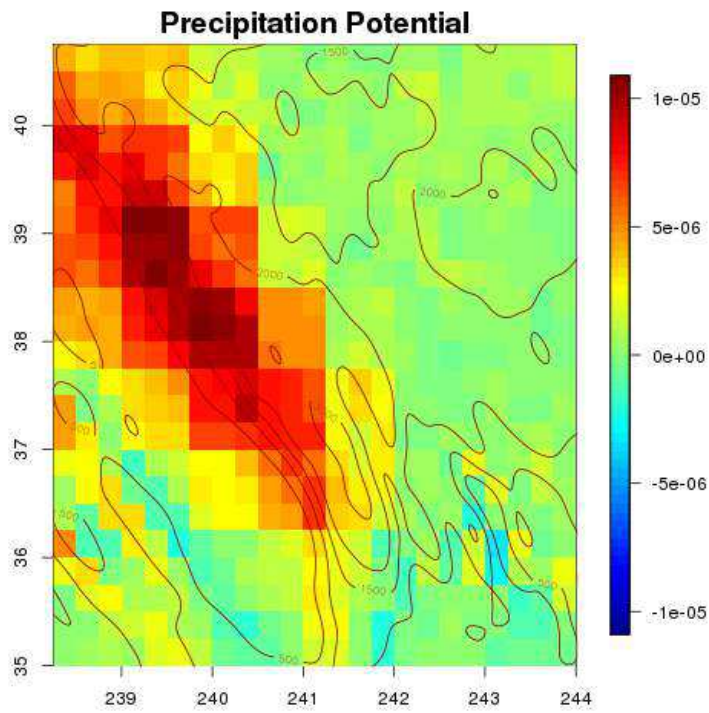
Fig. 3.35: Case Study on unevenly spaced data: 2 December 2005: Simulated precipitation potential
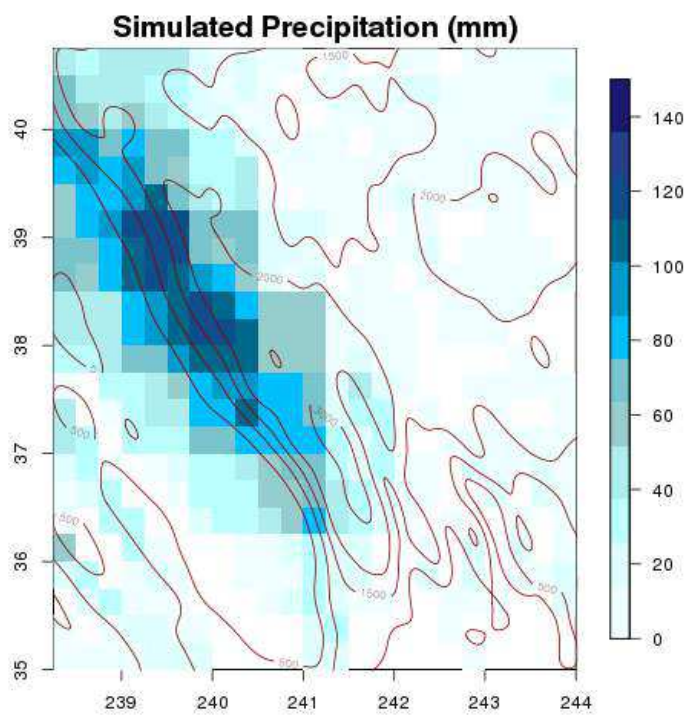
Fig. 3.36: Case Study on unevenly spaced data: 2 December 2005: Simulated precipitation amount

3.2.5 Case Study 4: Comparison with other methods

Comparison with Smith's Model

The first comparison is done in the 24hr spanning from 12:00 UTC Dec,1 2005 to 12:00 UTC Dec,2 2005. Simulation form Smith's model is done over the same grid proper of the CPC dataset used in all the previous experiments. The microphysical time scales were set $\tau_c = \tau_f = 2000s$, which sum is close to the MCMC estimates for $\tau = 4551$ found in the case study A on the same date. The background rate of precipitation LS was added to the source S following Smith and Barstad, 2004. Results are shown in Fig. 3.37
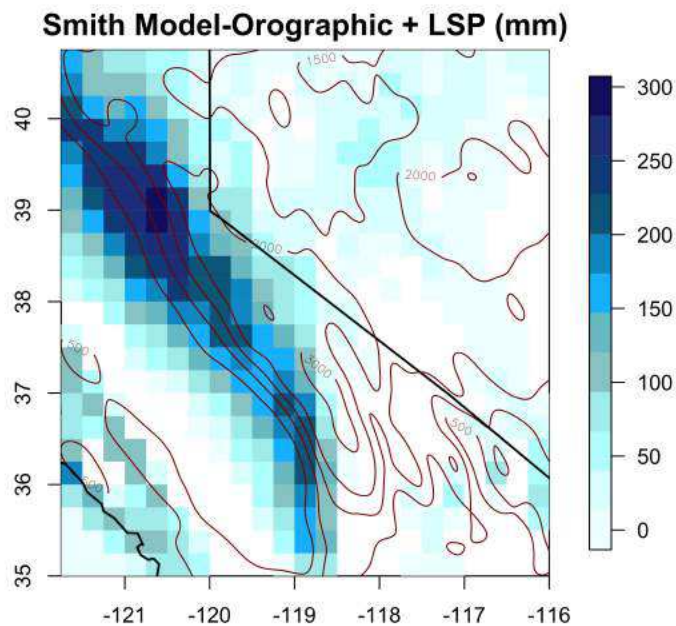
Fig. 3.37: Smith's Model, 2 December 2005: Simulated precipitation amount

Comparison with observations (CPC data) are shown in the following figure Fig. 3.38 for both Smith's model. The comparison is computed in relative differences $\frac{simulated-observed}{observed}$. Measure of skill indicates for this experiment a RMSE = 60.0542
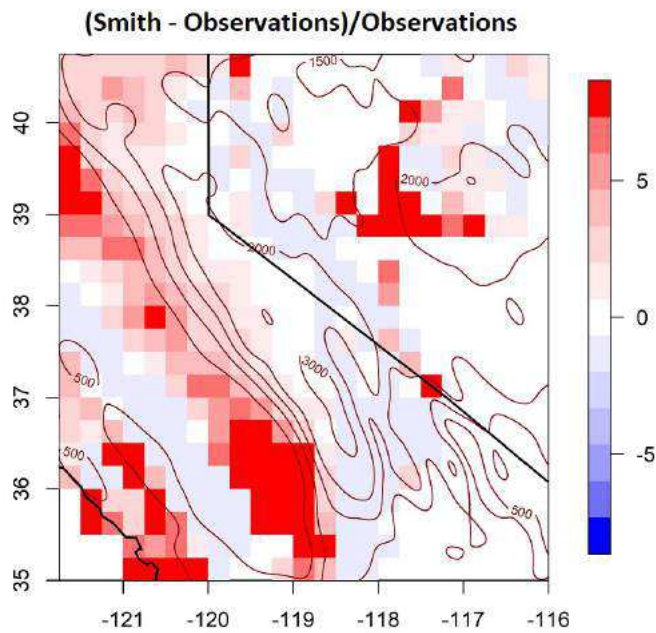
Fig. 3.38: Smith's Model: 2 December 2005: comparison with observations (CPC)

Comparison with Thin-Plate Smoothing Splines

The second comparison is done with respect to a statistical method of the kind proposed by Hutchinson, 1998. The analysis performed here tries to be as close as possible to the method described in the paper. Time interval: the 24hr spanning from 12:00 UTC Dec,1 2005 to 12:00 UTC Dec,2 2005; observed precipitation is daily accumulated precipitation from the Global Historical Climatology Network (GHCN) dataset. The analysis is done on the square root of precipitation amounts. The smoothing parameter is chosen in order to minimize the Generalized Cross Validation and independent variables used are Longitude, Latitude and Elevation.

Measures of skill indicated a RMSE = 18.01436 on the training set, and a pattern correlation PC = 0.851. The skill is comparable with that of the statistical model derived in this thesis when applied to gridded data. Thin Plate smoothing splines have instead a better skill when using unequally spaced station data.

Fig. 3.39 shows the prediction on the CPC dataset grid, achieved using the Thin Plate Smoothing Splines trained on the unequally spaced data. At the same Thin Plate smoothing splines do not provide indications on the processes leading to the observed precipitation pattern. Furthermore, as this method is a pure statistical interpolation methodology it cannot provide sub-grid signal with respect to a numerical model of interest.

Sim: Thin Plate Smoothing Splines
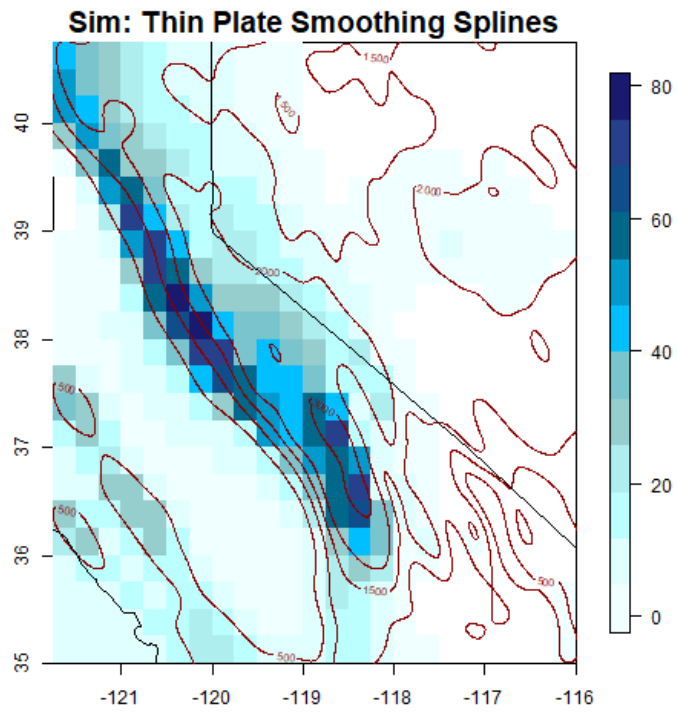
Fig. 3.39: Hutchinson's Thin Plate Smoothing Splines: 2 December 2005:

3.3 Figures and Tables

3.3.1 List of the experiments:
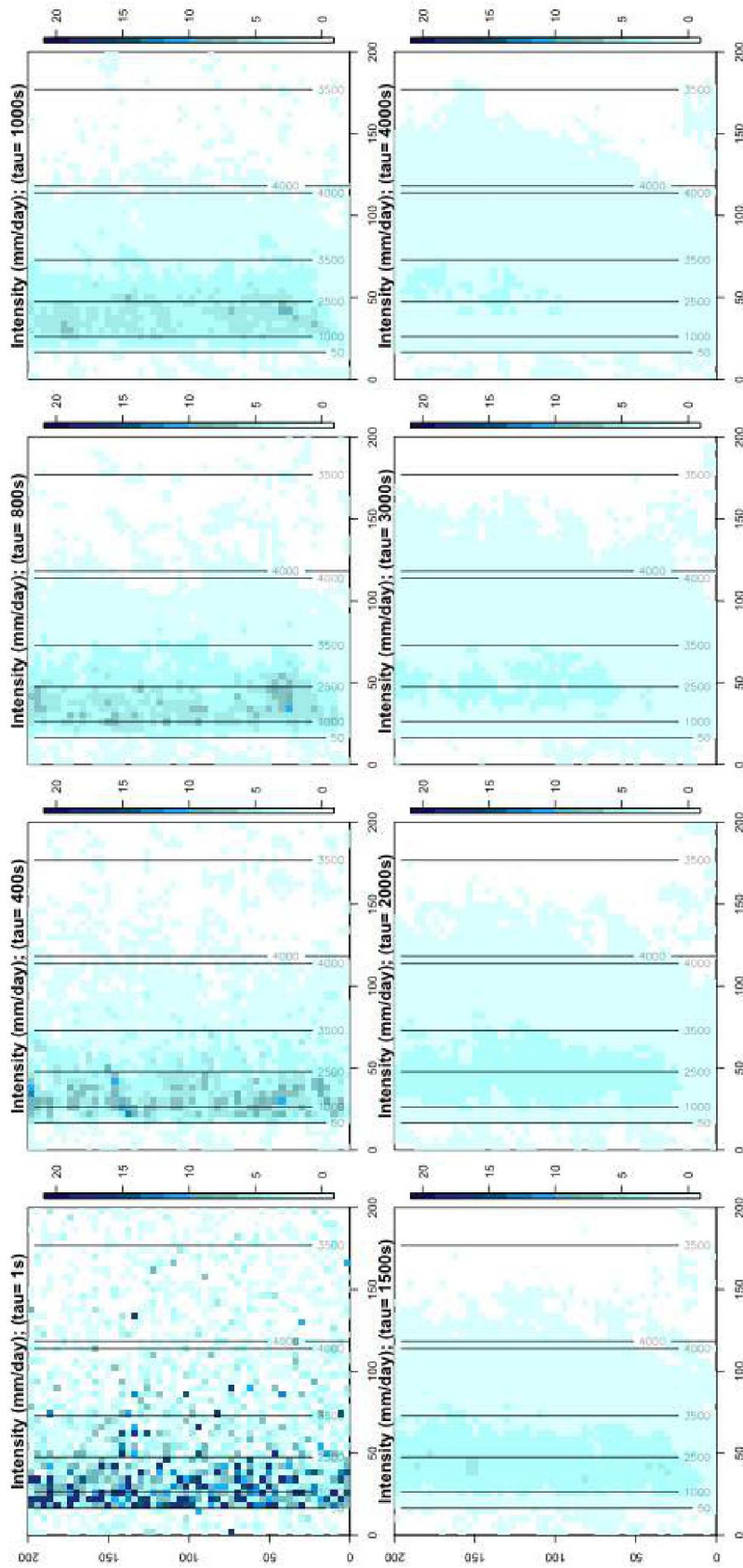
Fig. 3.40: North-South oriented mountain ridge - steep windward slope - SubExperiment A
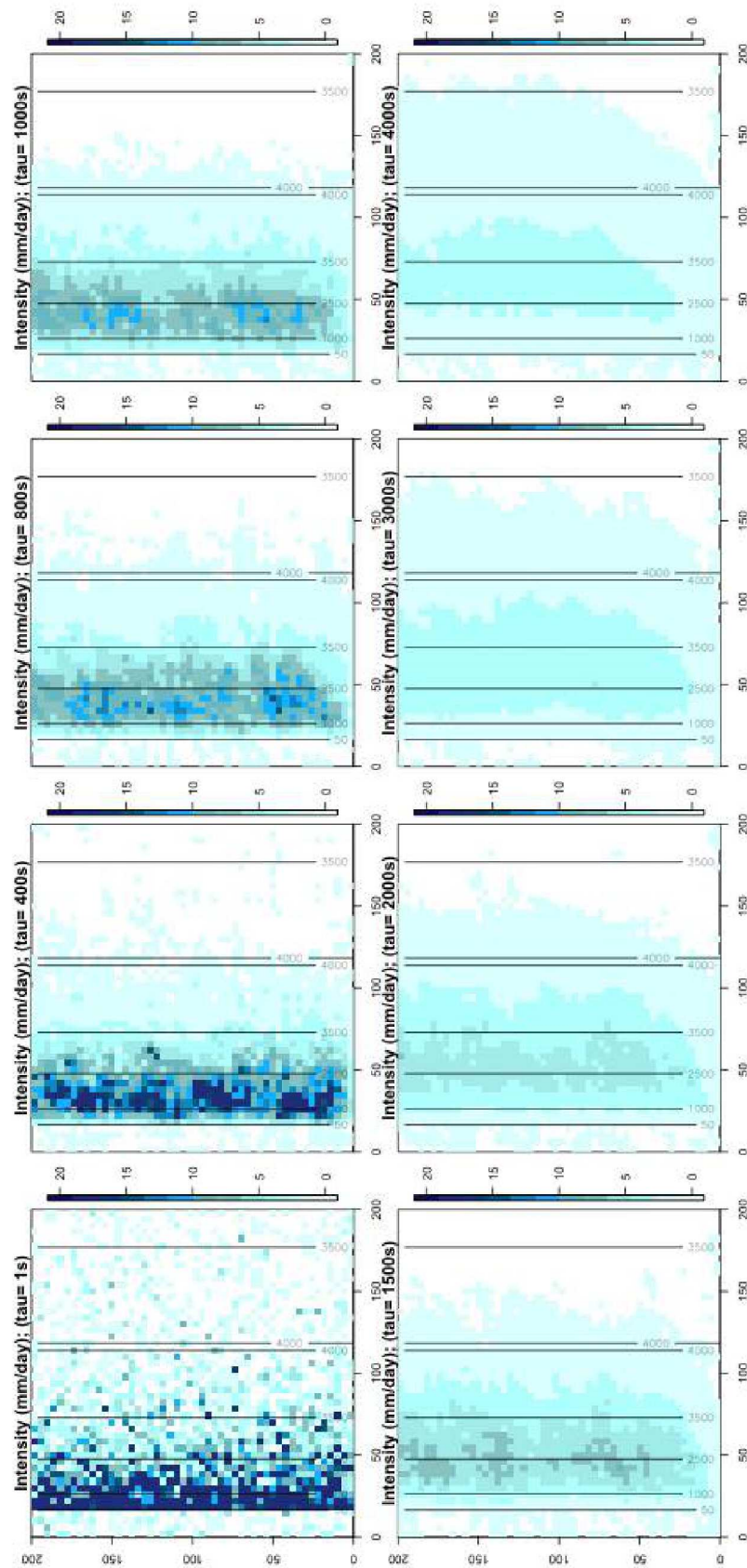
Fig. 3.41: North-South oriented mountain ridge - steep windward slope - SubExperiment B
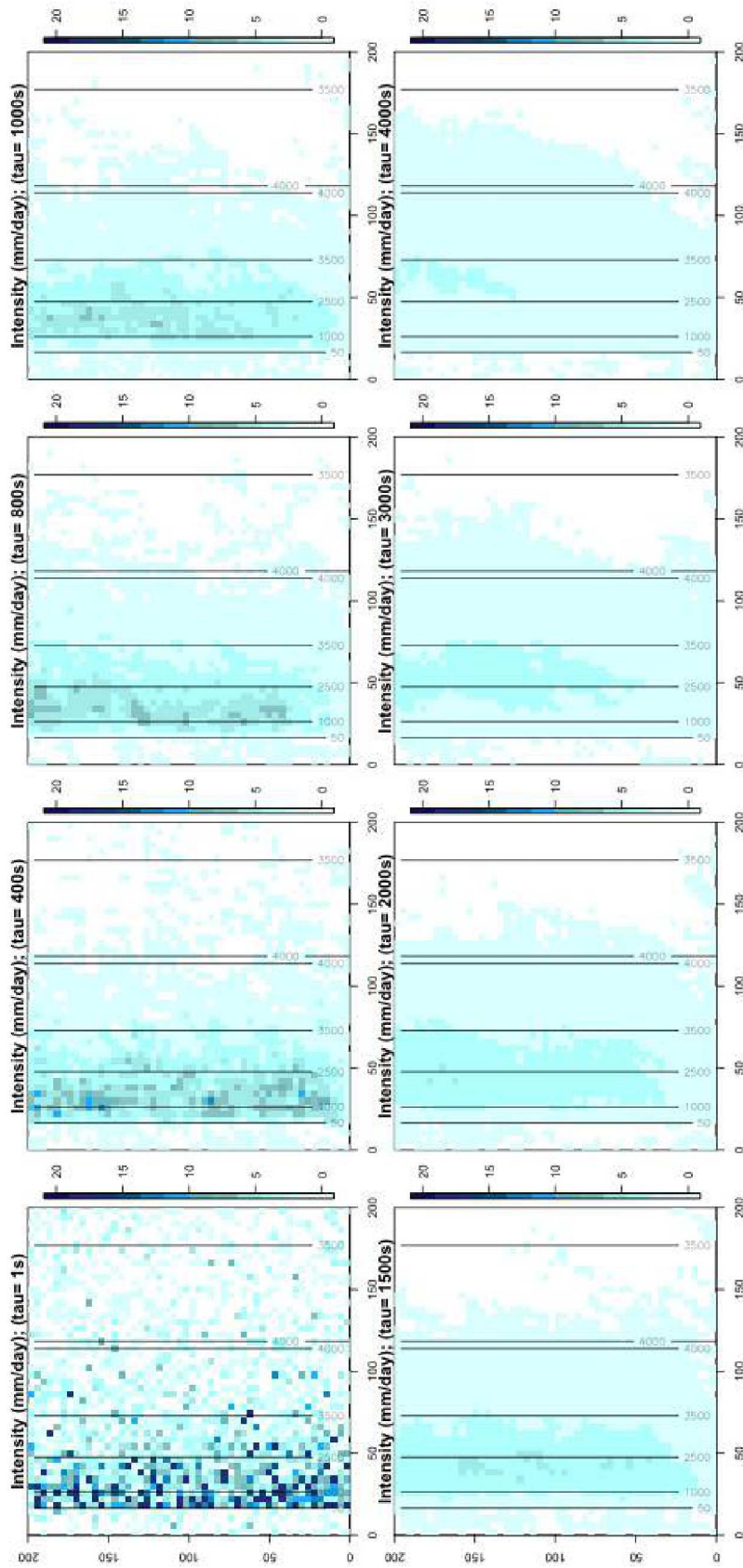
Fig. 3.42: North-South oriented mountain ridge - steep windward slope - SubExperiment C
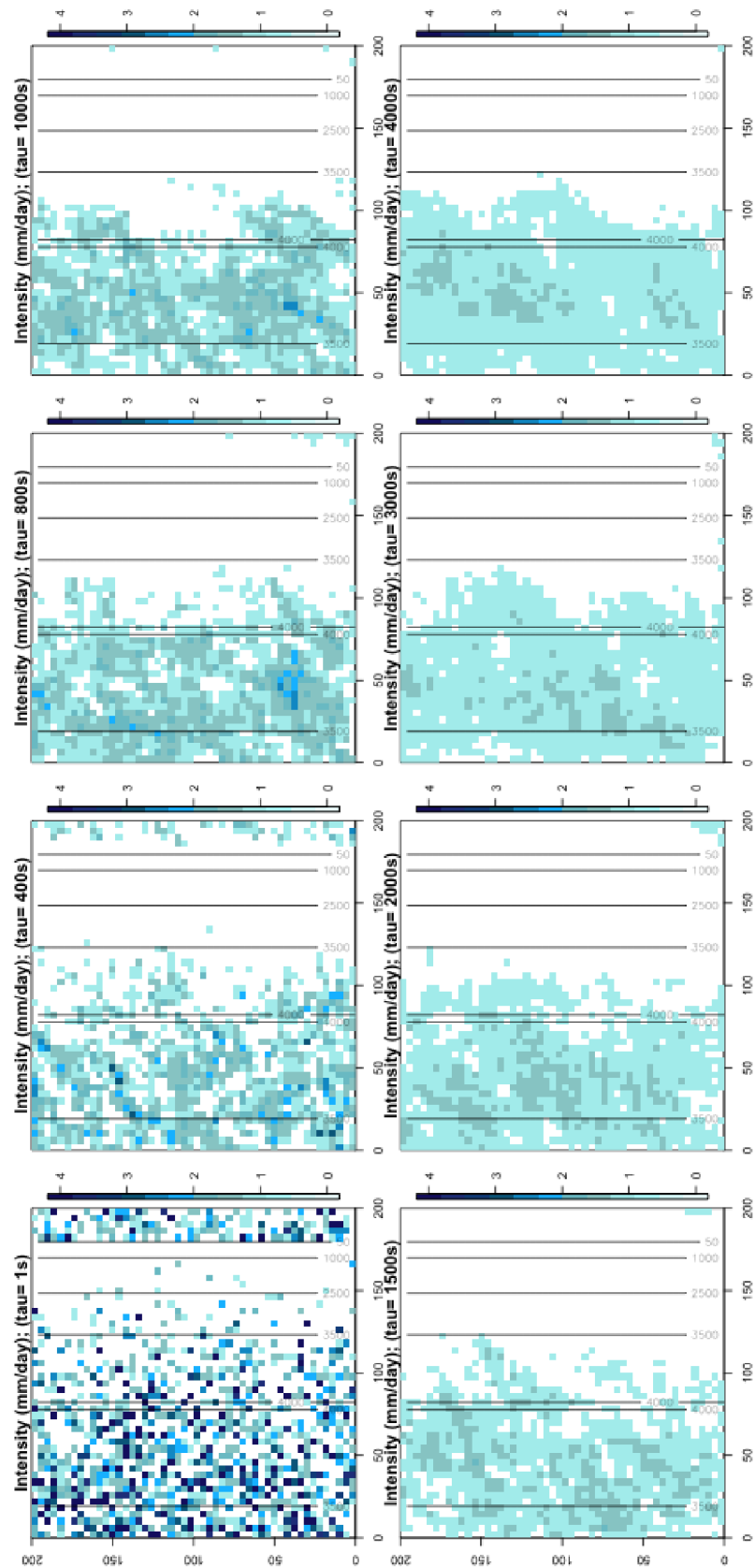
Fig. 3.43: North-South oriented mountain ridge - not-steep windward slope - SubExperiment A
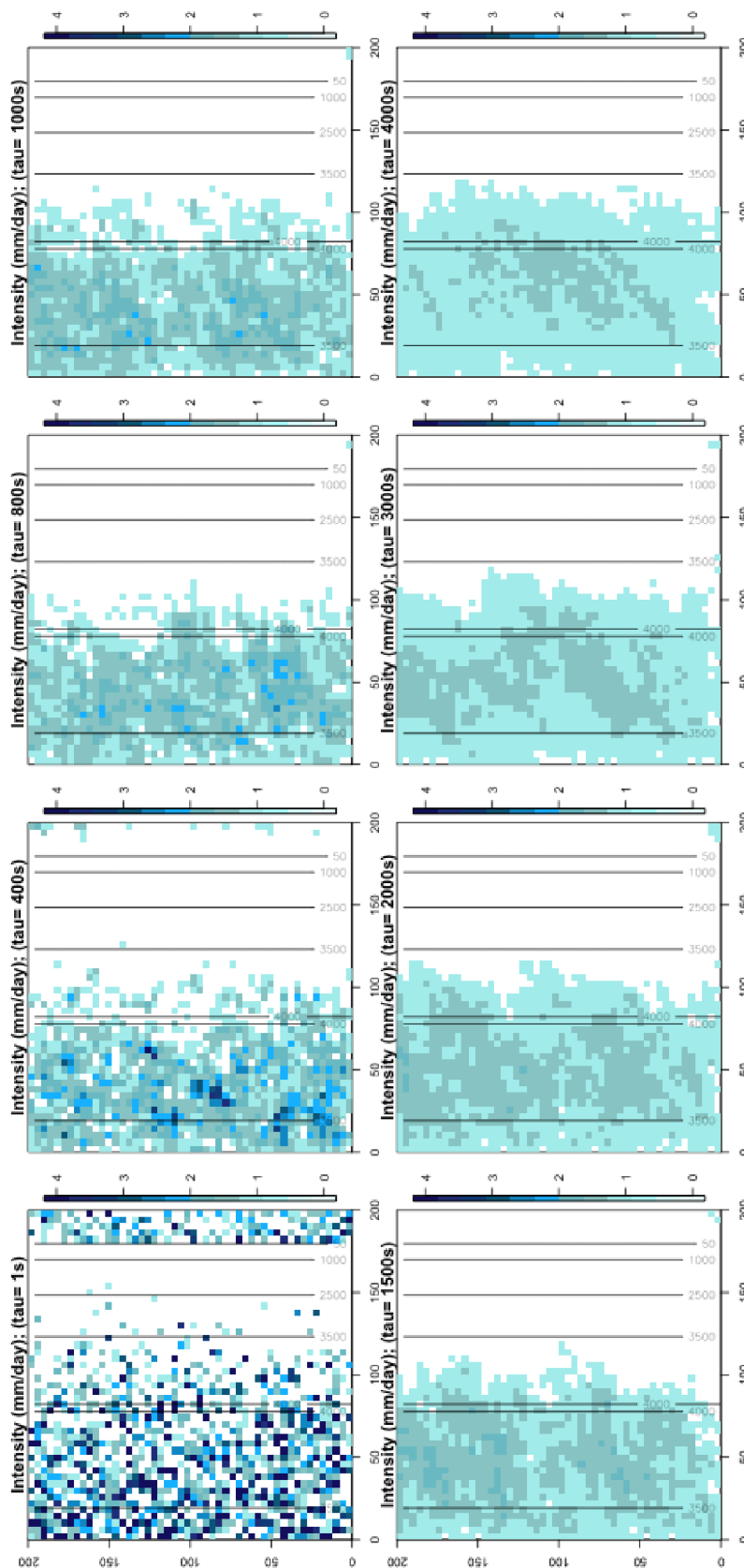
Fig. 3.44: North-South oriented mountain ridge - not-steep windward slope - SubExperiment B
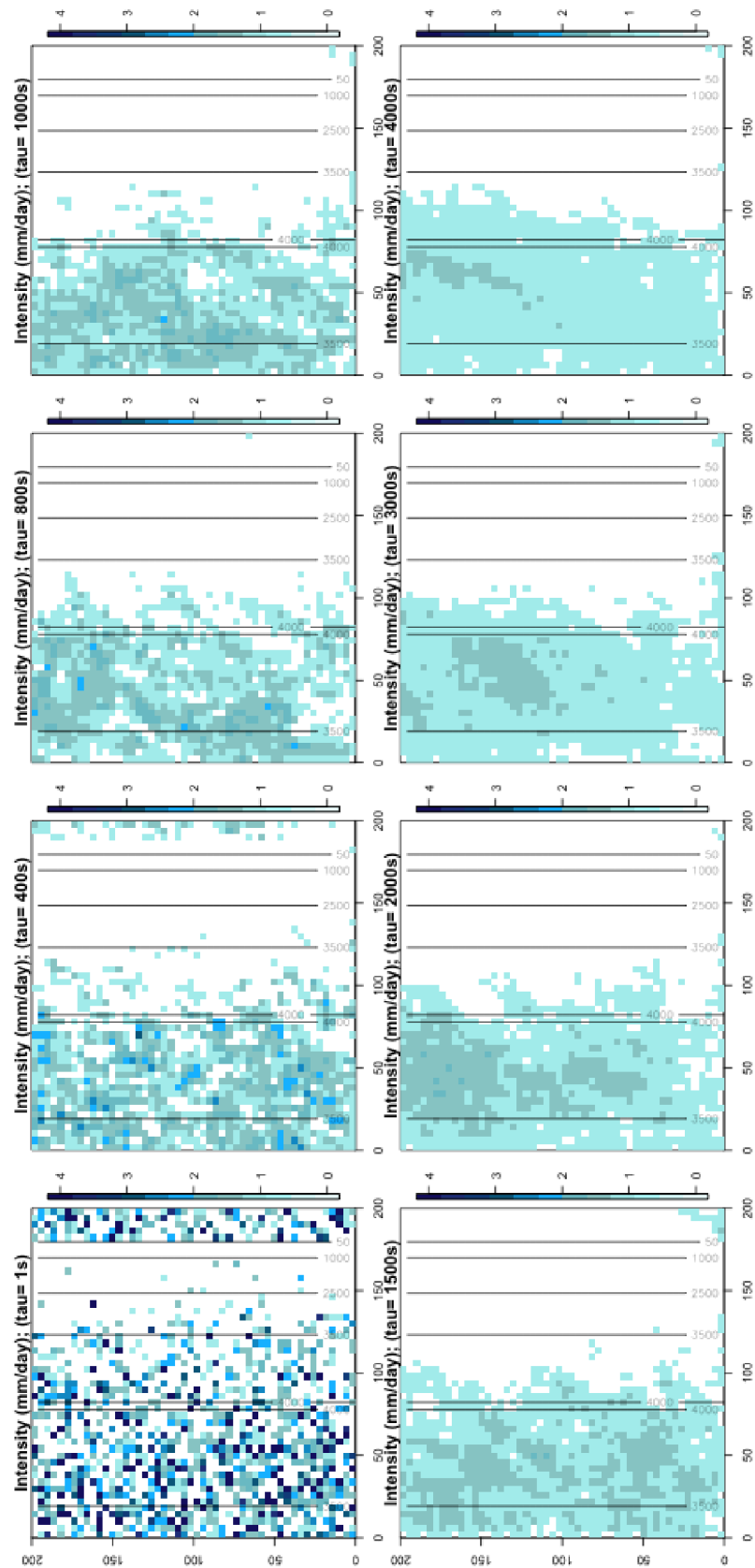
Fig. 3.45: North-South oriented mountain ridge - not-steep windward slope - SubExperiment C
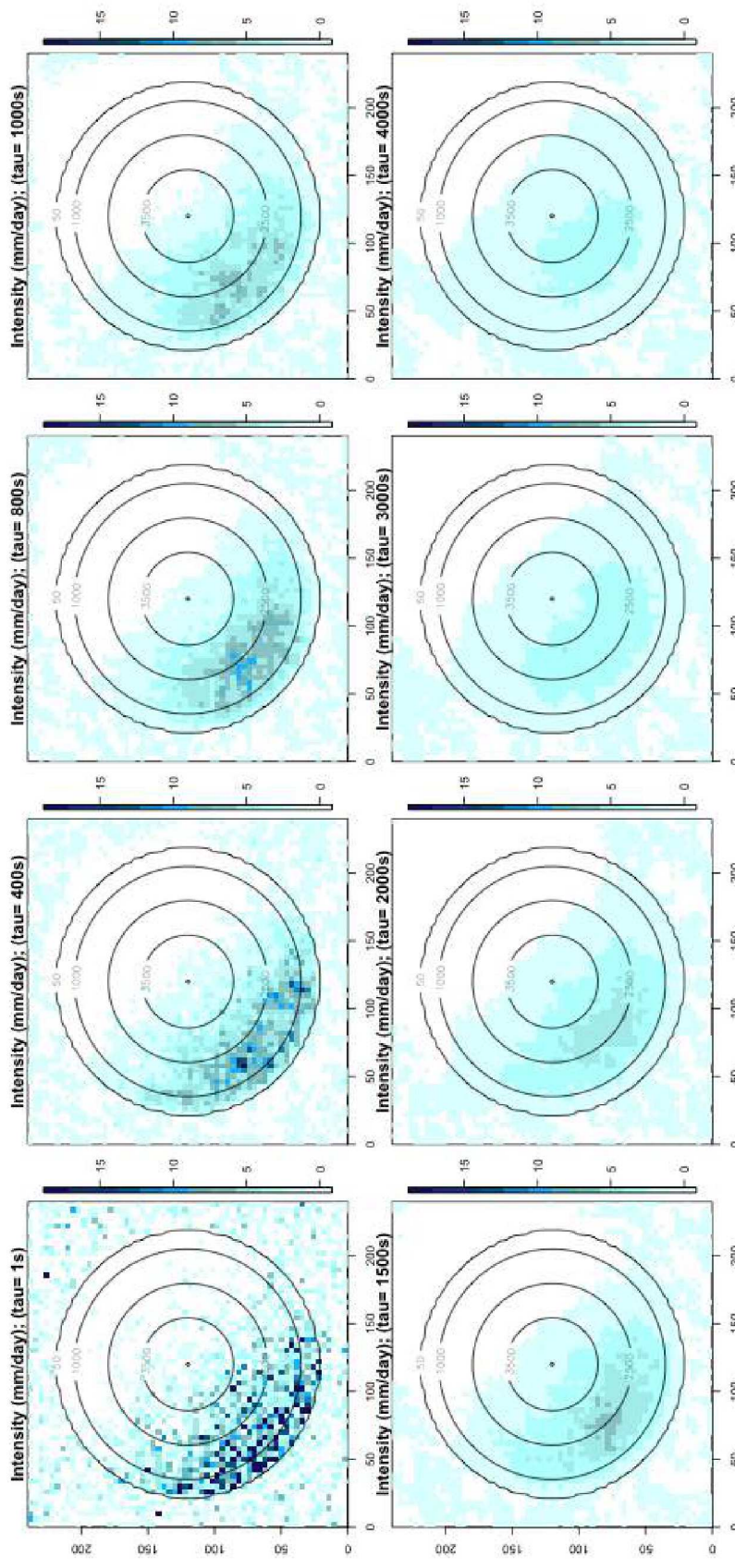
Fig. 3.46: Isolated Gaussian Ridge - SubExperiment A
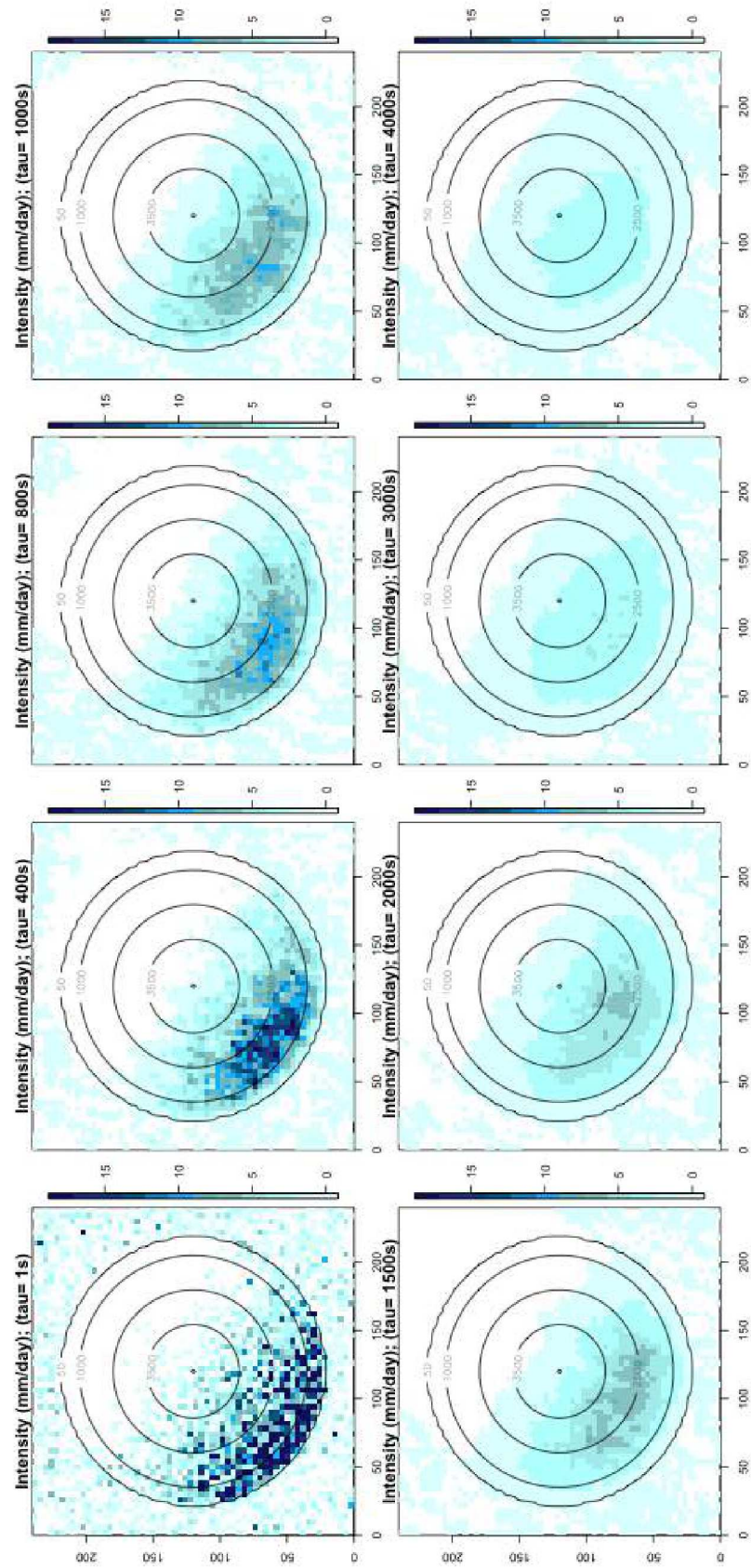
Fig. 3.47: Isolated Gaussian Ridge - SubExperiment B
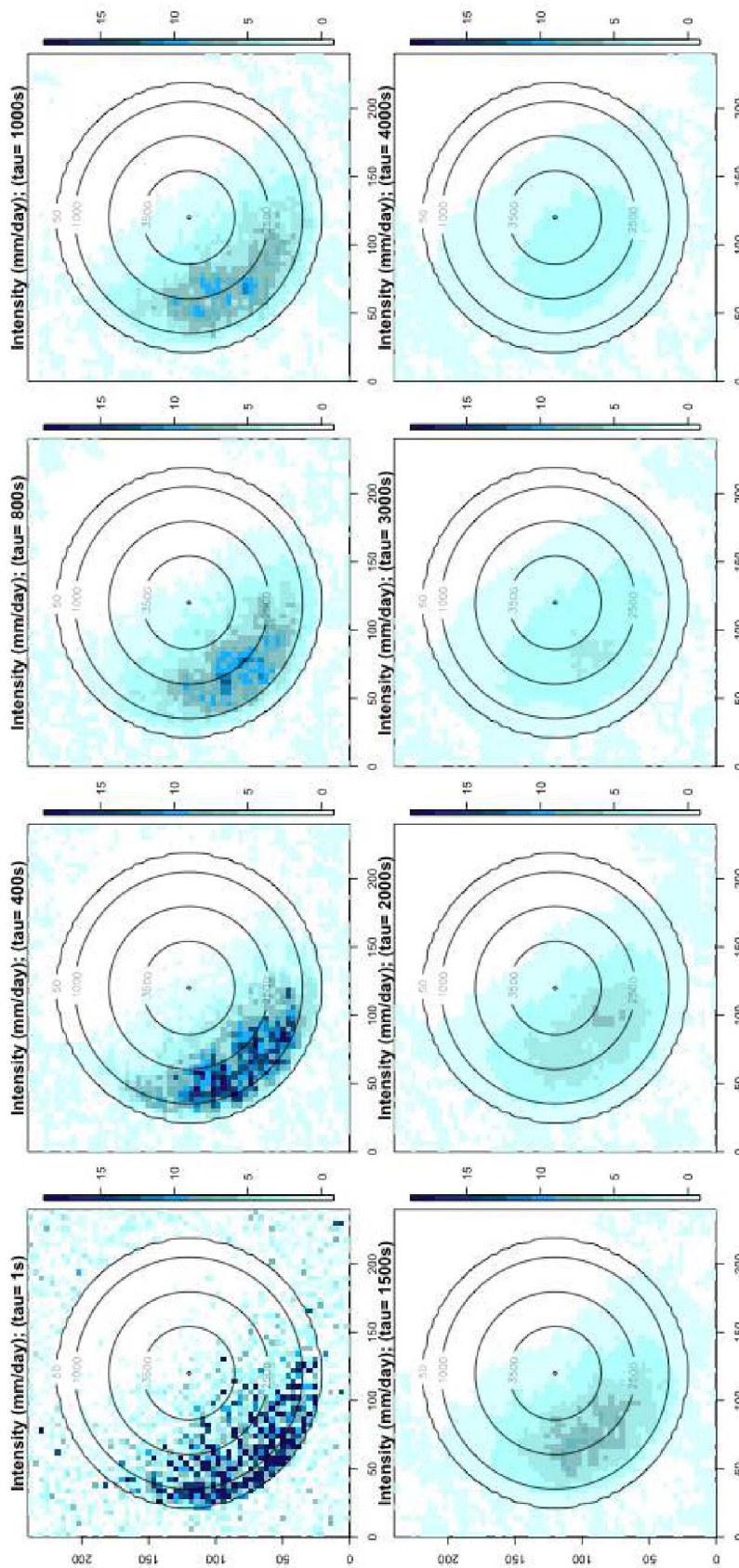
Fig. 3.48: Isolated Gaussian Ridge - SubExperiment C
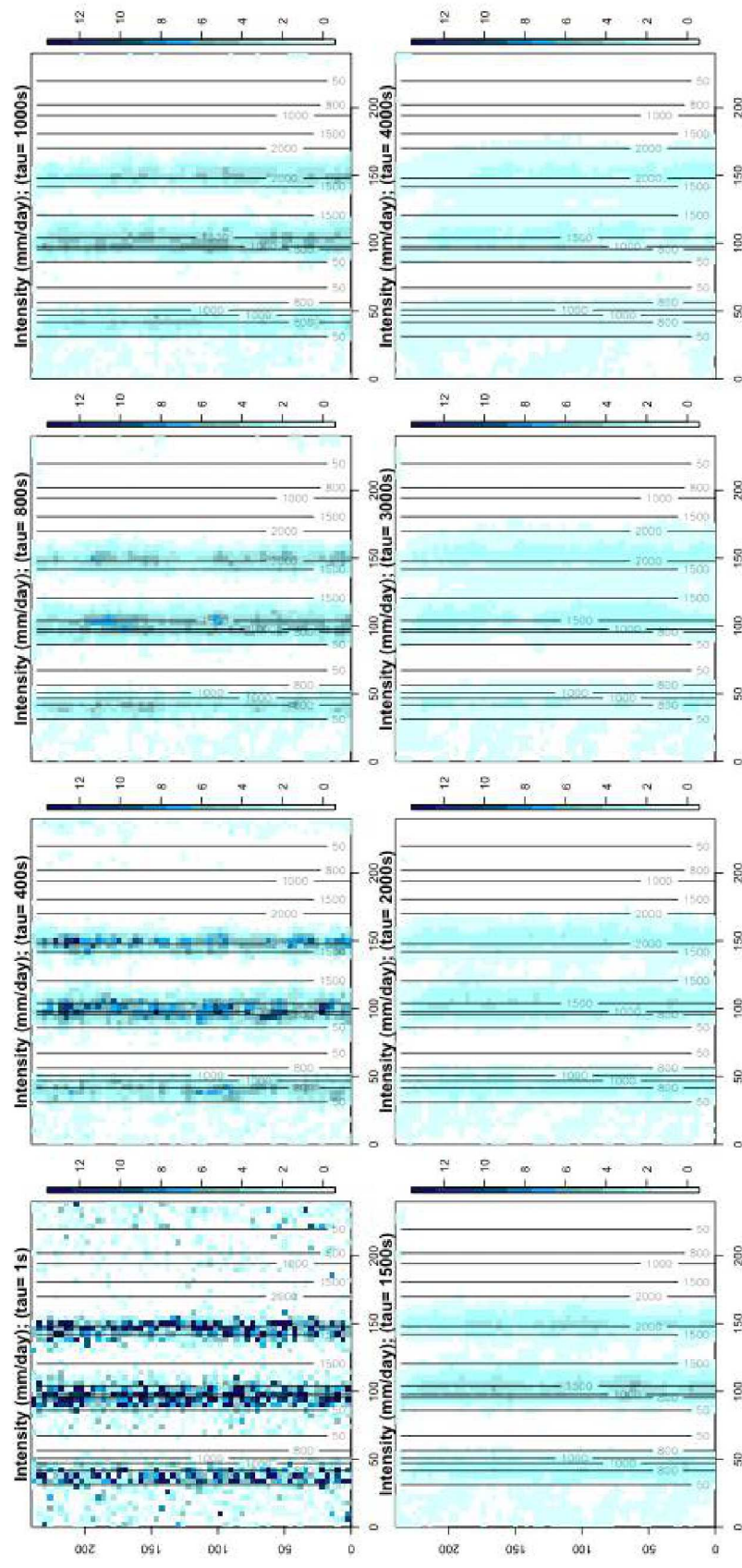
Fig. 3.49: Sequential Ridges - SubExperiment A
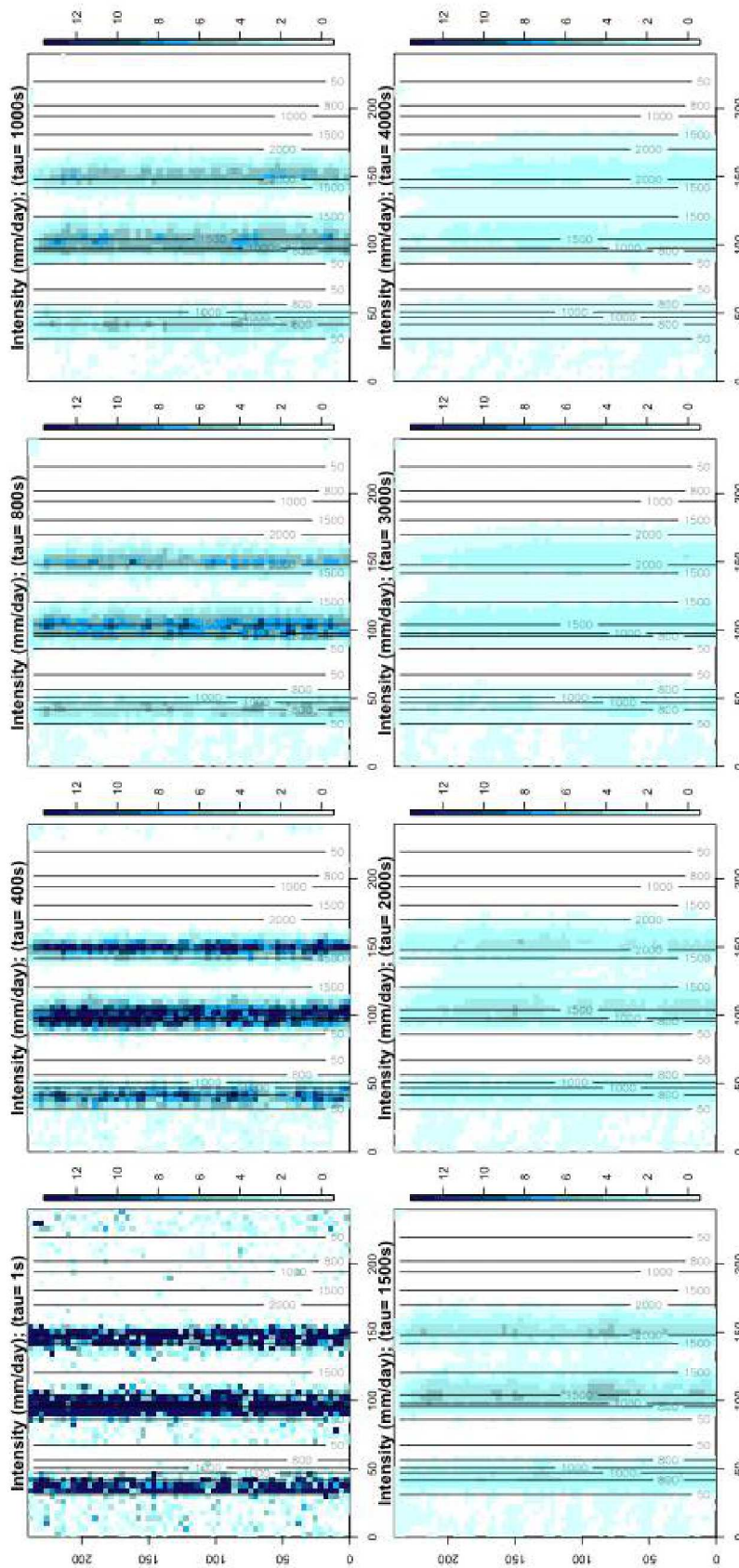
Fig. 3.50: Sequential Ridges - SubExperiment B
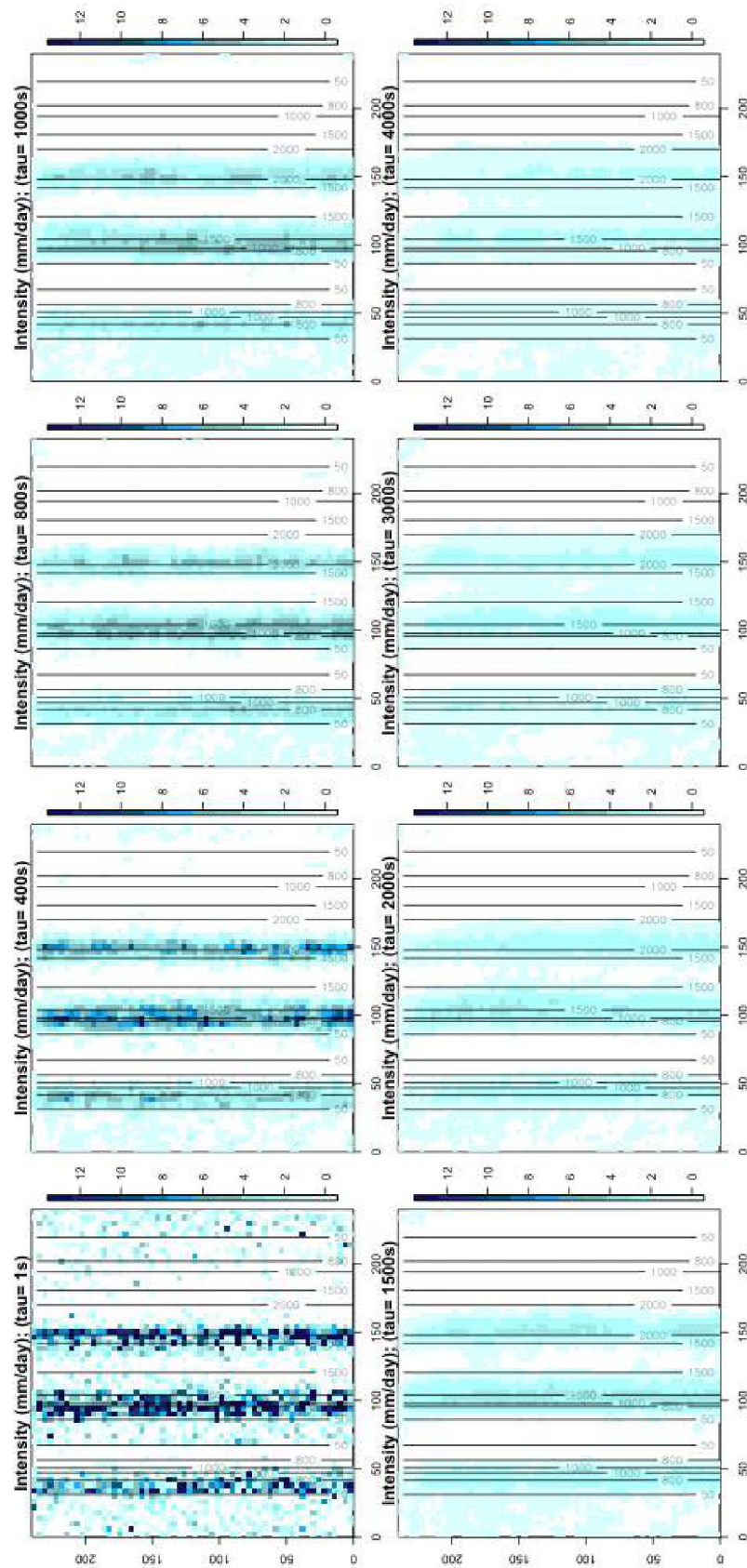
Fig. 3.51: Sequential Ridges - SubExperiment C

Discussion and Conclusion

<div style="text-align:right">4</div>

4.1 Discussion

The model 2.11 descends from a reduced version of Smith's model. The modification simplifies the original system of equations which model the advection of cloud water and hydrometeors as well as transition rates from water vapor to condensed water, from condensed water to *hydrometeors* and then to water fallout. Modifying the system to a single equation for cloud water density only (as shown in section 2.2), the precise description of multiple transition phases is lost, and only the transitions from water vapor to condensed water and then water fallout is made explicit. Because the transient phase indicated as "hydrometeor" is not usually a recordable quantity in observational campaigns, and because the two transitions occur in a sequential manner, the proposed simplification does not reduce the information in typical applications for which the statistical model was defined. Moreover, the reduction to a single equation allowed its discrete differences to define a valid Simultaneous Autoregressive Model, which theoretical properties are known, and allowed to build a more flexible framework.

The flexibility derives from the possibility for the atmospheric fields used (e.g. wind, moisture) to potentially vary in space. This is true for both the source function of condensation S (eq. 2.3) and for the advection of the condensed water. The same flexibility is given to the value of the time delay τ. Because the wind velocity and the parameter τ enter in the definition of both the mean and the covariance of the latent field p^*, the flexibility also allows a more adaptive definition of the moments of the process.

Since the model 2.11 identifies a valid SAR model, the knowledge on the physical processes included in the advection equation is combined with the information carried in observed data by means of unknown parameters estimation. Observations of precipitation are used to estimate the value of the parameter τ. τ is of fundamental importance, since it controls the time required to the cloud microphysical processes to occurr, it indirectly modulates the areas where precipitation falls on the ground, and finally precipitation intensity. Another advantage dealing with a SAR model is that, within its valid structure, many extensions and modifications are possible, making the model adaptive to different applications.

One example is the extension 2.33 presented in section 2.3, where precipitation generated in large-scale processes is included in the model definition. Information

carried in data is used to estimate τ and a coefficients of proportionality α, λ that modulates the effect of large-scale and local-scale precipitation on the process p^*. The model 2.33 assumes the microphysical time scale τ is equivalent in orographic processes and in large-scale processes.

A stochastic noise term was added to the discretized advection equation (section 2.2). The noise captures the uncertanty not explained in the analytical deterministic up-slope model and acts as further random impulses to the signal p^*. In principle, no restrictions are posed on the shape of the stochastic perturbation, if external knowledge on its spatial structure is available. The point to outline here is that, even adding a homoskedastic term, the resulting field p^* is heteroskedastic, and the non homogeneity in space is controlled by the atmospheric parameters u, v, the distance among the points in the lattice Δ_x, Δ_y and by the time scale τ.

Both models 2.11 and 2.33 are defined on a regular lattice. This might be a limitation in real applications, where observations are from sparse rain gauges. This restriction can be partially removed adapting the model for irregularly spaced data. One possible approach was presented in section 2.2.3. in which a criterium is given for the choice of neighbor points of each site.

Recalling the physical processes present in model 2.11, it is clear that only vertical motions due to uplift forced ascent over the mountain slope are included. This means that strictly convective local processes are not directly represented, although convection is most often triggered by an initial perturbation in the flow, as the one experienced in forced ascent over a topographic obstacle. Synoptic-scale convective storms can be included in the extension 2.33 but the daily cycle is not accounted. As already stated in chapter 2, the static stability, included in the source term S, enhances condensation, making larger, ahead upwind the mountain range, the area over which significant condensation is triggered. Stability does not have any effect on the flow dynamics: the horizontal wind is not affected, and it is thus not deflected when it impinges the range in presence of blocking conditions. This assumption is reasonable in the case of elongated ridges, but it is more questionable for circular mountain features or isolated peaks.

Once the value of parameter τ is estimated (and α or λ for the extension 2.33), the model becomes the governing process for p^* over the study area. Prediction is thus possible over un-sampled areas in which the same process can be similarly assumed valid, or over the same study area on a different grid.

The idealized experiments outlined the main features of the model in three different realistic yet simplified topographic features (elongated ridge, isolated peak, sequential ridges). The rate of condensation modeled in the source term S is proportional to the vertical wind velocity: steeper slopes or stronger cross-barrier wind increases the production of cloud water. Cloud water is then advected solely according to the speed of the horizontal wind. These features are clearly visible in the experiments

on an elongated ranges and in a more subtle way revealed in the experiments on the isolated peak. In the latter experiment, any increase in wind speed increases the condensation rate accordingly, regardless the direction of the wind, since the angle of the flow w.r.t the mountain slope is constant. The difference in angle is reflected solely in the locations where condensation occurs and in how cloud water is advected downwind.

A known drawback in Smith's model is the overestimation of condensation and precipitation rates over multiple ridges (Barstad and Smith, 2005). Indeed, the model does not have memory of precipitation occurred upwind in the flow, with the consequence that every windward slope acts as an independent obstacle to pass and the overall amount of condensed water might exceed the incoming moisture flux. This drawback is still present in model 2.11, as the experiment on idealized multiple ridges shows. The importance of the overestimation is indirectly reduced in the extension 2.33 of the model.

Chapter 3 reported real case studies, carried out over the Coastal Range and Sierra Nevada in California, USA, in winter days characterized by a Southwesterly wind. The spatial domain and the synoptic conditions represent a typical scenario in which orographic precipitation features are most evident. A case study on real data revealed how important is the inclusion in the framework of precipitation formed in large scale processes as well as its modulation by means of a coefficient of proportionality estimated from observations. By defining a customized index of accuracy, the results from the model shows a good agreement with observation. In the case study the atmospheric variables (e.g. wind, moisture, large-scale precipitation) were taken from ERA-Interim reanalysis. The possibility of using fields simulated in reanalysis or by climate models makes the statistical framework suitable to refine in space the spatial distribution of precipitation as downscaling procedure of climate models in retrospective studies and hindcasts. Other case studies showed the ability of the model in reproducing diverse observed precipitation patterns or regimes, as well as compares diverse methods for parameter estimation, from empirical scheme to Bayesian methods. Case studies also shows the potential to use the statistical framework on unevenly space data for example after applying the simple approach proposed in section 2.2.

The series of case studies concludes with the comparison of the statistical framework in Eq. 2.33 with other methodologies. In particular, a comparison with the original up-slope model used as kernel to derive the statistical model Eq. 2.11 shows some potential advantages achieved by the statistical framework as well as the similarities kept in reproducing the processes causing orographic precipitation. A second comparison was made with respect to a pure statistical interpolation technique, namely Thin Plate Smoothing Splines. The latter method exhibits a greater skill with respect to the model Eq. 2.33 when both are trained on unevenly spaced data and a comparable skill when applied on a gridded observational dataset.

To note is that Thin Plate Smoothing Splines are not linked to numerical models, thus are not designed for providing downscaled precipitation fields from numerical climate models.

The sensibility of parameter τ depends on the spatial scale of observed precipitation. In particular, variations in its values create greater variations in the simulated field as the spatial resolution of observations increase. New experiments testing this property should be carried out in the future changing the size of the domain considered and the resolution of the lattice.

4.2 Conclusion

In this work a Simultaneous Autoregressive Model for mapping the spatial distribution and intensity of precipitation over mountanous terrain is defined. The model includes the features typical of orographic precipitation, being based on an analytical model of the physical processes involved. At the same time, information in observed data contributes in the estimate of unknown parameters of the model. The model has a structure which allows extensions and refinements without modifying its basic structures and properties, making it suitable to applications in real case studies. The model, tested over mountain ranges in California, US, shows good agreement with observations.

The hybrid physical - statistical nature of the framework constitutes a new approach in literature for modeling orographic precipitation. The combination of the two approaches creates a synergy that overcomes some limitation typical of each of the two individually. Furthermore, the framework makes use of atmospheric fields as wind, humidity, temperatures and large-scale precipitation fields from reanalysis or numerical climate models. Thus the framework is able to provide downscaled precipitation for hindcast studies or to provide realistic values for climate model initialization. The framework may furthermore be used as kernel to build predictive downscaling methodologies applicable to diverse time scales, from short-range weather forecasts, to seasonal predictions, up to climate change projections.

Bibliography

Agnew, Maureen D and Jean P Palutikof (2000). „GIS-based construction of baseline climatologies for the Mediterranean using terrain variables“. In: *Climate Research* 14.2, pp. 115–127 (cit. on p. 14).

Albert, James H and Siddhartha Chib (1993). „Bayesian analysis of binary and polychotomous response data“. In: *Journal of the American statistical Association* 88.422, pp. 669–679 (cit. on pp. 30, 31).

Allard, Denis and Marc Bourotte (2015). „Disaggregating daily precipitations into hourly values with a transformed censored latent Gaussian process“. In: *Stochastic environmental research and risk assessment* 29.2, pp. 453–462 (cit. on p. 32).

Allcroft, David J and Chris A Glasbey (2003). „A latent Gaussian Markov random-field model for spatiotemporal rainfall disaggregation“. In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 52.4, pp. 487–498 (cit. on p. 32).

Alpert, Pinhas and Haim Shafir (1989). „Mesoγ-scale distribution of orographic precipitation: Numerical study and comparison with precipitation derived from radar measurements“. In: *Journal of applied meteorology* 28.10, pp. 1105–1117 (cit. on pp. 15, 16).

Anselin, Luc and Raymond Florax (2012). *New directions in spatial econometrics*. Springer Science & Business Media (cit. on pp. 24, 26).

Bader, MJ and WT Roach (1977). „Orographic rainfall in warm sectors of depressions“. In: *Quarterly Journal of the Royal Meteorological Society* 103.436, pp. 269–280 (cit. on p. 15).

Banerjee, Sudipto, Bradley P Carlin, and Alan E Gelfand (2014). *Hierarchical modeling and analysis for spatial data*. Crc Press (cit. on p. 30).

Barros, Ana Paula and Dennis P Lettenmaier (1994). „Dynamic modeling of orographically induced precipitation“. In: *Reviews of geophysics* 32.3, pp. 265–284 (cit. on p. 37).

Barstad, Idar and Felix Schüller (2011). „An extension of Smith's linear theory of orographic precipitation: Introduction of vertical layers“. In: *Journal of the Atmospheric Sciences* 68.11, pp. 2695–2709 (cit. on p. 19).

Barstad, Idar and Ronald B Smith (2005). „Evaluation of an orographic precipitation model“. In: *Journal of Hydrometeorology* 6.1, pp. 85–99 (cit. on p. 91).

Besag, Julian (1974). „Spatial interaction and the statistical analysis of lattice systems“. In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 192–236 (cit. on p. 24).

Brook, D (1964). „On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbour systems". In: *Biometrika* 51.3/4, pp. 481–483 (cit. on p. 24).

Buytaert, Wouter, Rolando Celleri, Patrick Willems, Bert De Bievre, and Guido Wyseure (2006). „Spatial and temporal rainfall variability in mountainous areas: A case study from the south Ecuadorian Andes". In: *Journal of hydrology* 329.3, pp. 413–421 (cit. on p. 13).

Chen, Tao, Liliang Ren, Fei Yuan, et al. (2017). „Comparison of Spatial Interpolation Schemes for Rainfall Data and Application in Hydrological Modeling". In: *Water* 9.5, p. 342 (cit. on pp. 13, 14).

Colle, Brian A (2004). „Sensitivity of orographic precipitation to changing ambient conditions and terrain geometries: An idealized modeling perspective". In: *Journal of the atmospheric sciences* 61.5, pp. 588–606 (cit. on p. 7).

– (2008). „Two-dimensional idealized simulations of the impact of multiple windward ridges on orographic precipitation". In: *Journal of the Atmospheric Sciences* 65.2, pp. 509–523 (cit. on p. 46).

Colle, Brian A, Ronald B Smith, and Douglas A Wesley (2013). „Theory, observations, and predictions of orographic precipitation". In: *Mountain Weather Research and Forecasting*. Springer, pp. 291–344 (cit. on p. 7).

Collier, CG (1975). „A representation of the effects of topography on surface rainfall within moving baroclinic disturbances". In: *Quarterly Journal of the Royal Meteorological Society* 101.429, pp. 407–422 (cit. on p. 15).

Corradini, C, F Melone, and RE Smith (2000). „Modeling local infiltration for a two-layered soil under complex rainfall patterns". In: *Journal of Hydrology* 237.1, pp. 58–73 (cit. on p. 37).

Cressie, Noël AC (1993). „Statistics for spatial data". In: (cit. on pp. 10, 23, 26, 27, 29).

Daly, Christopher (2006). „Guidelines for assessing the suitability of spatial climate data sets". In: *International journal of climatology* 26.6, pp. 707–721 (cit. on p. 13).

Daly, Christopher, Ronald P Neilson, and Donald L Phillips (1994). „A statistical-topographic model for mapping climatological precipitation over mountainous terrain". In: *Journal of applied meteorology* 33.2, pp. 140–158 (cit. on pp. 11, 15).

Dee, Dick P, SM Uppala, AJ Simmons, et al. (2011). „The ERA-Interim reanalysis: Configuration and performance of the data assimilation system". In: *Quarterly Journal of the royal meteorological society* 137.656, pp. 553–597 (cit. on pp. 39, 48).

Dempster, Arthur P, Nan M Laird, and Donald B Rubin (1977). „Maximum likelihood from incomplete data via the EM algorithm". In: *Journal of the royal statistical society. Series B (methodological)*, pp. 1–38 (cit. on p. 29).

Di Piazza, A, F Lo Conti, Leonardo V Noto, F Viola, and G La Loggia (2011). „Comparative analysis of different techniques for spatial interpolation of rainfall data to create a serially complete monthly time series of precipitation for Sicily, Italy". In: *International Journal of Applied Earth Observation and Geoinformation* 13.3, pp. 396–408 (cit. on p. 14).

Dingman, S Lawrence, Diana M Seely-Reynolds, and Robert C Reynolds (1988). „Application of kriging to estimating mean annual precipitation in a region of orographic influence". In: *JAWRA Journal of the American Water Resources Association* 24.2, pp. 329–339 (cit. on p. 10).

Dirks, KN, JE Hay, CD Stow, and D Harris (1998). „High-resolution studies of rainfall on Norfolk Island: Part II: Interpolation of rainfall data". In: *Journal of Hydrology* 208.3, pp. 187–193 (cit. on p. 9).

Dunn, Peter K (2004). „Occurrence and quantity of precipitation can be modelled simultaneously". In: *International Journal of Climatology* 24.10, pp. 1231–1239 (cit. on p. 32).

Feidas, Haralambos, Athanasios Karagiannidis, Stavros Keppas, et al. (2014). „Modeling and mapping temperature and precipitation climate data in Greece using topographical and geographical parameters". In: *Theoretical and applied climatology* 118.1-2, pp. 133–146 (cit. on p. 14).

Fleming, Mark M (2004). „Techniques for estimating spatially dependent discrete choice models". In: *Advances in spatial econometrics*. Springer, pp. 145–168 (cit. on p. 30).

Frei, Christoph and Christoph Schär (1998). „A precipitation climatology of the Alps from high-resolution rain-gauge observations". In: *International Journal of climatology* 18.8, pp. 873–900 (cit. on p. 11).

Geweke, John (1991). „Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities". In: *Computing science and statistics: Proceedings of the 23rd symposium on the interface*. Interface Foundation of North America, Inc., Fairfax Station, VA, pp. 571–578 (cit. on pp. 31, 32).

Ghil, Michael, MR Allen, MD Dettinger, et al. (2002). „Advanced spectral methods for climatic time series". In: *Reviews of geophysics* 40.1 (cit. on p. 48).

Goodale, Christine L, John D Aber, and Scott V Ollinger (1998). „Mapping monthly precipitation, temperature, and solar radiation for Ireland with polynomial regression and a digital elevation model". In: *Climate Research* 10.1, pp. 35–49 (cit. on p. 13).

Goovaerts, Pierre (1997). *Geostatistics for natural resources evaluation*. Oxford University Press on Demand (cit. on p. 10).

– (1998). „Ordinary cokriging revisited". In: *Mathematical Geology* 30.1, pp. 21–42 (cit. on p. 12).

– (2000). „Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall". In: *Journal of hydrology* 228.1, pp. 113–129 (cit. on p. 12).

Gottardi, Frédéric, Charles Obled, Joël Gailhard, and Emmanuel Paquet (2012). „Statistical reanalysis of precipitation fields based on ground network data and weather patterns: Application over French mountains". In: *Journal of Hydrology* 432, pp. 154–167 (cit. on p. 14).

Groisman, Pavel Ya and David R Legates (1994). „The accuracy of United States precipitation data". In: *Bulletin of the American Meteorological Society* 75.2, pp. 215–227 (cit. on p. 11).

Guan, Huade, John L Wilson, and Oleg Makhnin (2005). „Geostatistical mapping of mountain precipitation incorporating autosearched effects of terrain and climatic characteristics". In: *Journal of Hydrometeorology* 6.6, pp. 1018–1031 (cit. on p. 14).

Hobbs, Peter V, Richard C Easter, and Alistair B Fraser (1973). „A theoretical study of the flow of air and fallout of solid precipitation over mountainous terrain: Part II. Microphysics". In: *Journal of the Atmospheric Sciences* 30.5, pp. 813–823 (cit. on p. 15).

Houze, Robert A (2012). „Orographic effects on precipitating clouds". In: *Reviews of Geophysics* 50.1 (cit. on pp. 1, 7, 11, 37).

Hutchinson, Michael F (1998). „Interpolation of rainfall data with thin plate smoothing splines. Part I: Two dimensional smoothing of data with short range correlation". In: *Journal of Geographic Information and Decision Analysis* 2.2, pp. 139–151 (cit. on pp. 15, 74).

Jiang, Qingfang and Ronald B Smith (2003). „Cloud timescales and orographic precipitation". In: *Journal of the atmospheric sciences* 60.13, pp. 1543–1559 (cit. on pp. 6, 18, 20).

Katz, Richard W (1977). „Precipitation as a chain-dependent process". In: *Journal of Applied Meteorology* 16.7, pp. 671–676 (cit. on p. 32).

Katz, Richard W and Marc B Parlange (1995). „Generalizations of Chain-Dependent Processes: Application to Hourly Precipitation". In: *Water Resources Research* 31.5, pp. 1331–1341 (cit. on pp. 32, 33, 56).

Kleiber, William, Richard W Katz, and Balaji Rajagopalan (2012). „Daily spatiotemporal precipitation simulation using latent and transformed Gaussian processes". In: *Water Resources Research* 48.1 (cit. on p. 32).

LeSage, James P (2000). „Bayesian estimation of limited dependent variable spatial autoregressive models". In: *Geographical Analysis* 32.1, pp. 19–35 (cit. on pp. 30, 31).

– (2008). „An introduction to spatial econometrics". In: *Revue d'économie industrielle* 3, pp. 19–44 (cit. on pp. 24, 29, 30).

Li, Jin, Andrew D Heap, Anna Potter, and James J Daniell (2011). „Application of machine learning methods to spatial interpolation of environmental variables". In: *Environmental Modelling & Software* 26.12, pp. 1647–1659 (cit. on p. 14).

Li, Yifang and Sujit K Ghosh (2015). „Efficient sampling methods for truncated multivariate normal and student-t distributions subject to linear inequality constraints". In: *Journal of Statistical Theory and Practice* 9.4, pp. 712–732 (cit. on pp. 30–32).

Lloyd, CD (2005). „Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain". In: *Journal of Hydrology* 308.1, pp. 128–150 (cit. on p. 13).

Lu, George Y and David W Wong (2008). „An adaptive inverse-distance weighting spatial interpolation technique". In: *Computers & Geosciences* 34.9, pp. 1044–1055 (cit. on p. 9).

Luo, Xian, Youpeng Xu, and Yi Shi (2011). „Comparison of interpolation methods for spatial precipitation under diverse orographic effects". In: *Geoinformatics, 2011 19th International Conference on*. IEEE, pp. 1–5 (cit. on p. 11).

Ly, S., C. Charles, and A. Degré (2011a). „Geostatistical interpolation of daily rainfall at catchment scale: the use of several variogram models in the Ourthe and Ambleve catchments, Belgium". In: *Hydrology and Earth System Sciences* 15.7, pp. 2259–2274 (cit. on p. 12).

Ly, Sarann, Catherine Charles, and Aurore Degre (2011b). „Geostatistical interpolation of daily rainfall at catchment scale: the use of several variogram models in the Ourthe and Ambleve catchments, Belgium". In: *Hydrology and Earth System Sciences* 15.7, pp. 2259–2274 (cit. on p. 10).

Matheron, G (1971). *The theory of regionalized variables and its applications: École National Supérieure des Mines* (cit. on p. 9).

McMillen, Daniel P (1992). „Probit with spatial autocorrelation". In: *Journal of Regional Science* 32.3, pp. 335–348 (cit. on pp. 29, 30).

Meersmans, J, K Van Weverberg, S De Baets, et al. (2016). „Mapping mean total annual precipitation in Belgium, by investigating the scale of topographic control at the regional scale". In: *Journal of Hydrology* 540, pp. 96–105 (cit. on p. 14).

Neiman, Paul J, F Martin Ralph, AB White, DE Kingsmill, and POG Persson (2002). „The statistical relationship between upslope flow and rainfall in California's coastal mountains: Observations during CALJET". In: *Monthly Weather Review* 130.6, pp. 1468–1492 (cit. on pp. 39, 48).

Ninyerola, Miquel, Xavier Pons, and Joan M Roure (2007). „Monthly precipitation mapping of the Iberian Peninsula using spatial interpolation tools implemented in a Geographic Information System". In: *Theoretical and Applied Climatology* 89.3, pp. 195–209 (cit. on p. 14).

Phillips, Donald L, Jayne Dolph, and Danny Marks (1992). „A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain". In: *Agricultural and Forest Meteorology* 58.1-2, pp. 119–141 (cit. on p. 10).

Rhea, J Owen (1977). „Orographic precipitation model for hydrometeorological use". In: *Atmospheric science paper; no. 287* (cit. on p. 15).

Ripley, Brian D (2005). *Spatial statistics*. Vol. 575. John Wiley & Sons (cit. on pp. 24, 26, 27, 29).

Roe, Gerard H (2005). „Orographic precipitation". In: *Annu. Rev. Earth Planet. Sci.* 33, pp. 645–671 (cit. on pp. 1, 7).

Schiemann, R., R. Erdin, M. Willi, et al. (2011). „Geostatistical radar-raingauge combination with nonparametric correlograms: methodological considerations and application in Switzerland". In: *Hydrology and Earth System Sciences* 15.5, pp. 1515–1536 (cit. on p. 12).

Şen, Zekai and Zeyad Habib (2000). „Spatial precipitation assessment with elevation by using point cumulative semivariogram technique". In: *Water resources management* 14.4, pp. 311–325 (cit. on p. 14).

Sigrist, Fabio, Hans R Künsch, and Werner A Stahel (2011). „An autoregressive spatio-temporal precipitation model". In: *Procedia Environmental Sciences* 3, pp. 2–7 (cit. on p. 15).

– (2012). „A dynamic nonstationary spatio-temporal model for short term prediction of precipitation". In: *The Annals of Applied Statistics*, pp. 1452–1477 (cit. on p. 32).

Sinclair, Mark R (1994). „A diagnostic model for estimating orographic precipitation". In: *Journal of applied meteorology* 33.10, pp. 1163–1175 (cit. on p. 15).

Smith, Ronald B (1979). „The influence of mountains on the atmosphere". In: *Advances in geophysics* 21, pp. 87–230 (cit. on pp. 1, 2, 15, 16).

Smith, Ronald B (2003). „A linear upslope-time-delay model for orographic precipitation". In: *Journal of hydrology* 282.1, pp. 2–9 (cit. on pp. 15–17, 19, 20).

Smith, Ronald B and Idar Barstad (2004). „A linear theory of orographic precipitation". In: *Journal of the Atmospheric Sciences* 61.12, pp. 1377–1391 (cit. on pp. iii, 16–18, 51).

Smolarkiewicz, Piotr K and Richard Rotunno (1989). „Low Froude number flow past three-dimensional obstacles. Part I: Baroclinically generated lee vortices". In: *Journal of the Atmospheric Sciences* 46.8, pp. 1154–1164 (cit. on p. 5).

Sumner, Graham N (1988). *Precipitation: process and analysis*. 551.577 S9. Wiley New York (cit. on p. 4).

Tabios, Guillermo Q and Jose D Salas (1985). „A comparative analysis of techniques for spatial interpolation of precipitation". In: *JAWRA Journal of the American Water Resources Association* 21.3, pp. 365–380 (cit. on pp. 9, 10).

Thiessen, Alfred H (1911). „Precipitation averages for large areas". In: *Monthly weather review* 39.7, pp. 1082–1089 (cit. on p. 8).

Tobin, Cara, Ludovico Nicotina, Marc B Parlange, Alexis Berne, and Andrea Rinaldo (2011). „Improved interpolation of meteorological forcings for hydrologic applications in a Swiss Alpine region". In: *Journal of Hydrology* 401.1, pp. 77–89 (cit. on p. 13).

Vapnik, Vladimir Naumovich and Vlamimir Vapnik (1998). *Statistical learning theory*. Vol. 1. Wiley New York (cit. on p. 14).

Verbeek, Marno (2008). *A guide to modern econometrics*. John Wiley & Sons (cit. on p. 29).

White, Gentry and Sujit K Ghosh (2009). „A stochastic neighborhood conditional autoregressive model for spatial data". In: *Computational statistics & data analysis* 53.8, pp. 3033–3046 (cit. on p. 34).

Whittle, Peter (1954). „On stationary processes in the plane". In: *Biometrika*, pp. 434–449 (cit. on pp. 24, 28).

Wilks, Daniel S (2011). *Statistical methods in the atmospheric sciences*. Vol. 100. Academic press (cit. on p. 32).

Wilks, Daniel S and Robert L Wilby (1999). „The weather generation game: a review of stochastic weather models". In: *Progress in physical geography* 23.3, pp. 329–357 (cit. on p. 32).

List of Figures

Estratto per riassunto della tesi di dottorato

Studente: Paola Marson   matricola: 956160

Dottorato: Scienze e Gestione dei Cambiamenti Climatici

Ciclo: XXX

Titolo della tesi : A Statistical Framework for Modeling the Spatial Distribution and Intensity of Orographic Precipitation

Abstract:

Precipitation in mountainous areas exhibits a highly variability. Observations are in general too coarse in space to be representative of the variations of precipitation fields. To describie the spatial distribution and intensity of precipitation over complex terrain, we specify a statistical model which relies on both observed data and established analytical approximations of the physical processes involved. Approximations of local scale processes causing orographic precipitation are combined with a precipitation component due to large scale processes in an autoregressive framework. Observation are used to adjust the contribution of the two sources.
The statistical model inherently uses atmospheric fields provided by numerical climate models. Thus, the framework can be nested into climate models and provide a process-informed and spatially-consistent sub-grid refinement adjusted by observations.

Abstract:

La precipitazione osservata in aree montuose è caratterizzata da una elevata variabilità spesso non rappresentata nelle osservazioni disponibili. Allo scopo di descrivere la distribuzione spaziale e l'intensità

della precipitazione in aree montuose in questa tesi viene definito un modello statistico basato sia sulle osservazioni disponibile che su descrizione dei processi che causano precipitazione orografica. Processi di piccola scala e la componente di larga scala della precipitazione sono unite in uno schema autoregressivo. Osservazioni sono usate per pesare opportunamente il contributo delle due fonti. Il modello statistico usa campi atmosferici generati da modelli climatici numerici, quindi puo' essere innestato in modelli climatici e generare segnale sub-grid sulla base dei processi di piccola scala e ulteriormente corretto da osservazioni disponibili.