

Bases de Datos de Radiación Solar

Energía Solar Fotovoltaica

Oscar Perpiñán Lamigueiro

Universidad Politécnica de Madrid

① Bases de Datos

② Control de Calidad

① Bases de Datos

Introducción

Estaciones Meteorológicas

Estaciones Meteorológicas: modelos empíricos

Imágenes de Satélite

Fuentes de Datos: Estaciones Terrestres

Fuentes de Datos: Satélite

Métodos híbridos

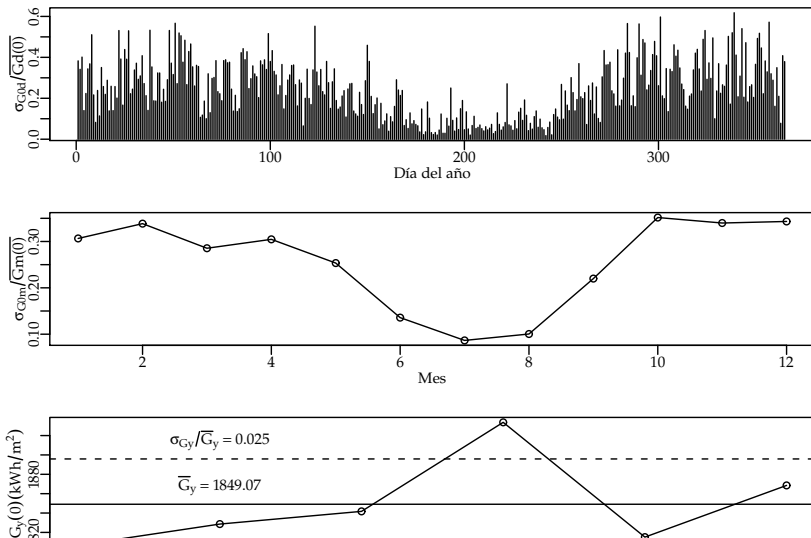
② Control de Calidad

Variabilidad Temporal y Espacial

- ▶ La irradiancia solar extraterrestre depende de la latitud y el instante temporal (*proceso determinista*).
- ▶ La irradiancia solar incidente en la superficie terrestre es resultado de la interacción con la atmósfera cambiante: **variabilidad temporal y espacial** (*proceso estocástico*).

Variabilidad Temporal

Variabilidad de la irradiación diaria, mensual y anual durante el período comprendido entre 2001-2008 en Carmona, Sevilla



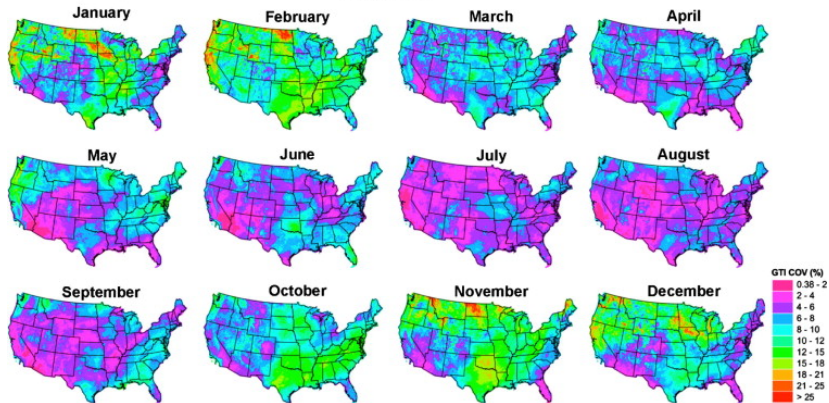
Variabilidad Temporal

$$\sigma_{\overline{G}} = \frac{\sigma_G}{\sqrt{N}}$$

- ▶ Predicción para un (día, mes, año) **determinado**:
 - ▶ Intervalo de confianza del 95 % acotado por $1,96 \cdot \sigma_G$
- ▶ Predicción para un (día, mes, año) **promedio (durante N años)**:
 - ▶ Intervalo de confianza del 95 % acotado por $1,96 \cdot \sigma_{\overline{G}}$

Variabilidad Espacial

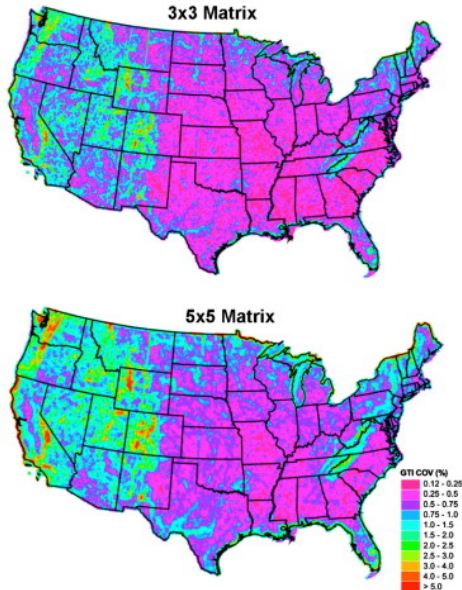
Monthly GTI Interannual COV (%)
1998-2005



$$COV = 1/G_p \sqrt{\frac{\sum_1^n (G_p^2 - G_i^2)}{n}}$$

Variabilidad Espacial

Spatial GTI COV (%) of Annual Average 1998-2005



Estimación a partir de Medidas

- ▶ Para estimar la radiación incidente es necesario contar con:
 - ▶ **Medidas cercanas** (variabilidad espacial): distancia no superior a 10 km.
 - ▶ **Series temporales** largas (variabilidad temporal): 10 años.

Fuentes de datos

► Estaciones meteorológicas

- Series largas y con tiempos de muestreo altos.
- Baja resolución espacial (medidas puntuales)
- Precisión en caso de medida directa.
- Tipos:
 - Con medidor de radiación
 - Sin medidor de radiación (modelos empíricos).

Fuentes de datos

▶ Estaciones meteorológicas

- ▶ Series largas y con tiempos de muestreo altos.
- ▶ Baja resolución espacial (medidas puntuales)
- ▶ Precisión en caso de medida directa.
- ▶ Tipos:
 - ▶ Con medidor de radiación
 - ▶ Sin medidor de radiación (modelos empíricos).

▶ Imágenes de satélite

- ▶ Tiempos de muestreo bajos (mejorando)
- ▶ Resolución espacial alta
- ▶ Error debido a la estimación.

Fuentes de datos

▶ Estaciones meteorológicas

- ▶ Series largas y con tiempos de muestreo altos.
- ▶ Baja resolución espacial (medidas puntuales)
- ▶ Precisión en caso de medida directa.
- ▶ Tipos:
 - ▶ Con medidor de radiación
 - ▶ Sin medidor de radiación (modelos empíricos).

▶ Imágenes de satélite

- ▶ Tiempos de muestreo bajos (mejorando)
- ▶ Resolución espacial alta
- ▶ Error debido a la estimación.

▶ Híbrido

- ▶ Medidas terrestres combinadas con imágenes de satélite

① Bases de Datos

Introducción

Estaciones Meteorológicas

Estaciones Meteorológicas: modelos empíricos

Imágenes de Satélite

Fuentes de Datos: Estaciones Terrestres

Fuentes de Datos: Satélite

Métodos híbridos

② Control de Calidad

Estaciones Meteorológicas: medida directa

La medida directa de radiación solar se realiza con un piranómetro.



- ▶ Pila termoeléctrica (termopares con barniz negro)
- ▶ Alojamiento con dos hemiesferas de cristal.
- ▶ Flujo de calor por radiación provoca tensión eléctrica en termopila.

Estaciones Meteorológicas: medida directa

La medida directa de radiación solar se realiza con un piranómetro.



- ▶ Respuesta espectral plana para radiación visible.
- ▶ Respuesta perfecta al coseno del ángulo de incidencia (pérdidas por reflexión).

Estaciones Meteorológicas: medida directa

La medida directa de radiación solar se realiza con un piranómetro.

- ▶ Requiere mantenimiento y calibración frecuente.

La red de estaciones que miden directamente radiación es escasa para estimaciones precisas en regiones grandes

- ▶ La proporción de estaciones con piranómetros es baja respecto a las que miden temperatura ambiente y precipitación (1:500).

① Bases de Datos

Introducción

Estaciones Meteorológicas

Estaciones Meteorológicas: modelos empíricos

Imágenes de Satélite

Fuentes de Datos: Estaciones Terrestres

Fuentes de Datos: Satélite

Métodos híbridos

② Control de Calidad

Frente a la baja densidad de estaciones con medida directa de radiación se emplean modelos empíricos

- ▶ Relaciones entre radiación y otras variables
 - ▶ Horas de brillo (*sunshine duration*)
 - ▶ Cobertura nubosa
 - ▶ Temperatura ambiente
 - ▶ Precipitación
 - ▶ Humedad
 - ▶ ...
- ▶ Los coeficientes de los modelos sólo se pueden ajustar en estaciones con medidas de radiación.
- ▶ Los coeficientes dependen del lugar de ajuste, pero se pueden interpolar para otras localizaciones.

Estaciones Meteorológicas: modelos empíricos

- ▶ Radiación y Horas de Brillo (Angstrom y Prescott)

$$\frac{G(0)}{B_o(0)} = a_1 + b_1 \frac{S}{S_o}$$

- ▶ Problema: poca disponibilidad de datos

Estaciones Meteorológicas: modelos empíricos

- ▶ Radiación y Temperatura (Bristow y Campbell)

$$G(0) = a (1 - \exp(-b\Delta T^c)) \cdot B_o(0)$$

- ▶ Variaciones con más variables: Lluvia (si/no), rango antes y después, velocidad viento, humedad relativa.

$$G(0) = a (1 - \exp(-b\Delta T^c)) \cdot B_o(0) \cdot \left(1 + \sum_1^n p_j \cdot v_j \right) + p_{n+1}$$

① Bases de Datos

Introducción

Estaciones Meteorológicas

Estaciones Meteorológicas: modelos empíricos

Imágenes de Satélite

Fuentes de Datos: Estaciones Terrestres

Fuentes de Datos: Satélite

Métodos híbridos

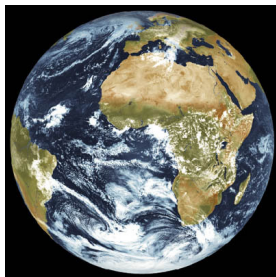
② Control de Calidad

Fundamentos

- ▶ Los satélites meteorológicos están equipados con **radiómetros** (sensores de radiación electromagnética a diferentes frecuencias) que captan **radiación emitida por la Tierra**.
- ▶ La radiación emitida por la Tierra depende de la **reflexión del suelo**, y la **geometría y composición de la atmósfera**.
- ▶ Diferentes fenómenos físicos se detectan en **bandas de frecuencias** distintas (canales).
- ▶ Existen diversos procedimientos para **estimar radiación solar** en superficie a partir de la información de los diferentes canales del radiómetro.

Satelites Geoestacionarios Europeos: Meteosat

- ▶ **MFG:** Meteosat First Generation (7 satélites)
 - ▶ Equipados con el radiómetro MVIRI (Meteosat Visible and Infrared Imager).
 - ▶ Tres canales: visible, infrarrojo, vapor de agua.
- ▶ **MSG:** Meteosat Second Generation (3 satélites)
 - ▶ Equipados con dos radiómetros:
 - ▶ **SEVIRI** (Spinning Enhanced Visible and InfraRed Imager): 12 canales
 - ▶ GERB (Geostationary Earth Radiation Budget): infrarrojo visible.



Procedimientos: Heliosat-2

Pasos

- ▶ Establecer **albedo de referencia** (*suelo*).
- ▶ Estimar **índice de cobertura nubosa**.
- ▶ Estimar radiación en superficie a partir de cobertura nubosa y **modelo de cielo claro**.

- ▶ Empleado para base HelioClim
- ▶ Usan datos de MVIRI
- ▶ Accesible via SoDa:
<http://www.soda-is.com/heliosat/index.html>

Procedimientos: CM SAF

► Fundamento:

- Se emplea un **Radiative Transfer Model (RTM)**, libRadtran, para generar una matriz de estados (**Look-up table, LUT**) relaciona la transmitancia atmosférica y el albedo de la atmósfera para variedad de estados.
- La irradiancia en superficie se estima multiplicando la irradiancia extra-atmosférica por la **transmitancia atmosférica determinada interpolando en la LUT**.

Procedimientos: CM SAF

► **Fundamento:**

- Se emplea un **Radiative Transfer Model (RTM)**, libRadtran, para generar una matriz de estados (**Look-up table, LUT**) relaciona la transmitancia atmosférica y el albedo de la atmósfera para variedad de estados.
- La irradiancia en superficie se estima multiplicando la irradiancia extra-atmosférica por la **transmitancia atmosférica determinada interpolando en la LUT**.

► **Dos LUTs:** cielo nuboso, cielo claro.

► **Cielo nuboso:**

- Estimación de albedo y estado atmosférico a partir de imágenes.
- Estimación de transmitancia interpolando en LUT para cielo nuboso.

► **Cielo claro:**

- Estimación de transmitancia interpolando en LUT para cielo claro **sin estimación previa** de albedo.

Procedimientos: CM SAF

► **Fundamento:**

- Se emplea un **Radiative Transfer Model (RTM)**, libRadtran, para generar una matriz de estados (**Look-up table, LUT**) relaciona la transmitancia atmosférica y el albedo de la atmósfera para variedad de estados.
- La irradiancia en superficie se estima multiplicando la irradiancia extra-atmosférica por la **transmitancia atmosférica determinada interpolando en la LUT**.

► **Dos LUTs:** cielo nuboso, cielo claro.

► **Cielo nuboso:**

- Estimación de albedo y estado atmosférico a partir de imágenes.
- Estimación de transmitancia interpolando en LUT para cielo nuboso.

► **Cielo claro:**

- Estimación de transmitancia interpolando en LUT para cielo claro **sin estimación previa** de albedo.

► Emplean datos del **radiómetro MSG/SEVIRI**

Procedimientos: LSA SAF

- ▶ Generación de **máscara de nubes** a partir de imagen usando algoritmo de **NWC-SAF**.
- ▶ Para **zonas sin nubes**: modelo de cielo claro sin usar datos de imagen.
- ▶ Para **zonas cubiertas**: modelo de transmitancia atmosférica a partir de imágenes.
- ▶ Emplean datos del **radiómetro MSG/SEVIRI**

① Bases de Datos

Introducción

Estaciones Meteorológicas

Estaciones Meteorológicas: modelos empíricos

Imágenes de Satélite

Fuentes de Datos: Estaciones Terrestres

Fuentes de Datos: Satélite

Métodos híbridos

② Control de Calidad

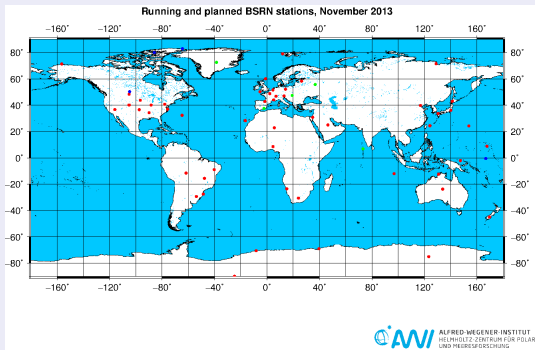
Wiki con recursos

<https://github.com/oscarperpinan/mds/wiki>

Baseline Surface Radiation Network

<http://www.bsrn.awi.de/>

- BSRN provides near-continuous, long-term, in situ-observed, Earth-surface, broadband irradiances (solar and thermal infrared) and certain related parameters from a network of more than 50 globally diverse sites.



Baseline Surface Radiation Network

- ▶ Validation and confirmation of satellite and computer model estimates.
- ▶ Datos desde: http://www.bsrn.awi.de/en/data/data_retrieval_via_pangaea/

Measurement and Instrumentation Data Center NREL

<http://www.nrel.gov/midc/>

Radiación global, directa y difusa (y otras variables) con muestreo de 1 min en diversas localidades de EEUU.



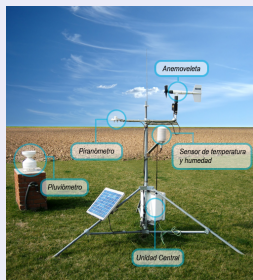
<https://eportal.mapa.gob.es/websiar/Inicio.aspx>

- ▶ El Sistema de Información Agroclimática para el Regadío (SiAR) registra datos agroclimáticos relacionados con demanda hídrica de las zonas de riego.
- ▶ Más de 400 estaciones.
- ▶ Valores diarios y horarios



Sensores

- ▶ Temperatura y Humedad
- ▶ Piranómetro
- ▶ Anemoveleta
- ▶ Pluviómetro
- ▶ Temperatura del suelo (algunas)



Radiación

- ▶ Alrededor de 30 estaciones en todo el territorio.
- ▶ Medidas de global, difusa y directa.
- ▶ Sólo gráficas.

Estaciones «convencionales»

- ▶ Presión, temperatura, viento, humedad, lluvia.
- ▶ Permite descarga de datos horarios por día.

Redes de Comunidades Autónomas

- ▶ Meteogalicia
- ▶ MeteoNavarra
- ▶ Cataluña
- ▶ MeteoEuskadi
- ▶ Andalucía

① Bases de Datos

Introducción

Estaciones Meteorológicas

Estaciones Meteorológicas: modelos empíricos

Imágenes de Satélite

Fuentes de Datos: Estaciones Terrestres

Fuentes de Datos: Satélite

Métodos híbridos

② Control de Calidad

Wiki con recursos

<https://github.com/oscarperpinan/mds/wiki>

Surface meteorology and Solar Energy (SSE)

- ▶ 200 satellite-derived meteorology and solar energy parameters **monthly averaged** from 22 years of data
- ▶ Resolución $1^{\circ} \times 1^{\circ}$

<https://eosweb.larc.nasa.gov/cgi-bin/sse/sse.cgi>

- ▶ **EUMETSAT** is the European operational satellite agency for monitoring weather, climate and the environment.
- ▶ **Satellite Application Facilities (SAFs)**
 - ▶ Dedicated centres of excellence for processing satellite data.
 - ▶ Generate and disseminate operational EUMETSAT products and services.

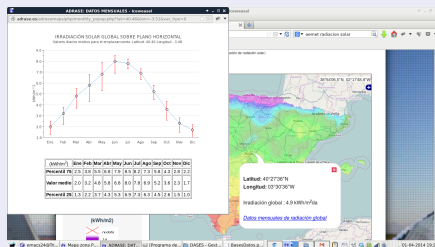
- ▶ **SAF on Climate Monitoring (CM SAF):** provision of satellite-derived geophysical parameter data sets suitable for **climate monitoring**
 - ▶ Environmental Data Records (EDR): time-tagged earth-located geophysical parameters produced from sensor data. EDRs are derived in low to medium latency not fulfilling strictest climate requirements.
 - ▶ Climate Data Records (CDR): time series of measurements of sufficient length, consistency, and continuity to determine climate variability and change.
- ▶ **SAF on Land Surface Analysis (LSA SAF):** generates, archives and disseminates, on an **operational basis**, a set of parameters involved in the surface radiation budget, evapotranspiration, vegetation cover and and fire-related products.

SAFs: Radiación

- ▶ **CM SAF:** Surface incoming shortwave radiation (**SIS**)
 - ▶ AEMET ha analizado las estimaciones para España en su [Atlas de Radiación](#).
- ▶ **LSA SAF:** Down-welling surface short-wave radiation flux (**DSSF**)

<http://adrase.es>

- ▶ Radiación solar media mensual, resolución aproximada de 5x5 km.
 - ▶ Media mensual y anual más probable durante un periodo de largo plazo (imágenes de satélite, modelo aproximadamente Heliosat)
 - ▶ Variabilidad esperada de los valores diarios mensuales: (series largas de datos de estaciones de AEMET y extrapolación espacial con IDW)



① Bases de Datos

Introducción

Estaciones Meteorológicas

Estaciones Meteorológicas: modelos empíricos

Imágenes de Satélite

Fuentes de Datos: Estaciones Terrestres

Fuentes de Datos: Satélite

Métodos híbridos

② Control de Calidad

Interpolación Espacial

Objetivo: mejorar la resolución espacial de medidas dispersas

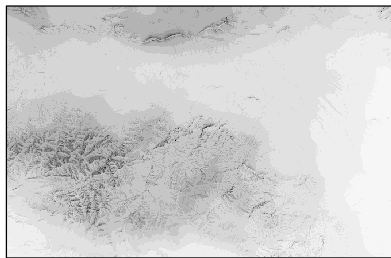
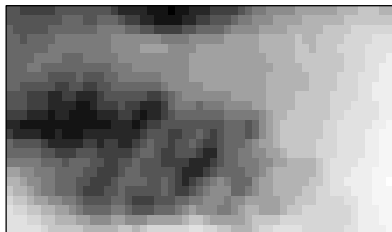
- ▶ **Inverse Distance Weighting (IDW)**: determinista.
- ▶ **Ordinary Kriging**: modelo determinista para la media (constante) y estocástico para residuos.

$$\hat{z}(\mathbf{s}) = \mu + \epsilon(\mathbf{s})$$

- ▶ **Kriging with External Drift (KED)**: modelo determinista para la media incorporando información de una variable con alta densidad espacial.

$$\hat{z}(\mathbf{s}_\theta) = \sum_{k=0}^p \hat{\beta}_k q_k(\mathbf{s}_\theta) + \sum_{i=1}^n \lambda_i \epsilon(\mathbf{s}_i)$$

Corrección por topografía



Sky-View Factor (SVF) Proporción de cielo visible para un receptor horizontal (afecta a la radiación difusa isotrópica)

$$SVF = 1 - \int_0^{2\pi} \sin^2 \theta_{hor} d\theta$$

Horizon blocking Bloqueo de región circumsolar por horizonte: afecta a radiación directa y difusa anisotrópica

<http://re.jrc.ec.europa.eu/pvgis/apps4/pvest.php>

PVGIS (Photovoltaic Geographical Information System) is a research, demonstration and policy-support instrument for geographical assessment of the solar energy resource in the context of integrated management of distributed energy generation.

- ▶ Computation of clear-sky global irradiation on a horizontal surface
- ▶ Sky obstruction by local terrain features (hills or mountains) calculated from the digital elevation model.
- ▶ Interpolation of the clear-sky index and computation of global irradiation on a horizontal surface.

① Bases de Datos

② Control de Calidad

① Bases de Datos

② Control de Calidad

Estadística

Gráficos

Control de Calidad de Medidas

Control de Calidad de Modelos

Variable aleatoria y proceso estocástico

- ▶ Una **variable aleatoria** es una función que asigna un único número real a cada resultado de un espacio muestral en un experimento.
- ▶ Un **proceso estocástico** es una variable aleatoria que evoluciona a lo largo del **tiempo** (p.ej. la radiación).

Función de densidad de probabilidad

La función de densidad de probabilidad, $f(X)$, de una variable aleatoria **asigna probabilidad** a un suceso:

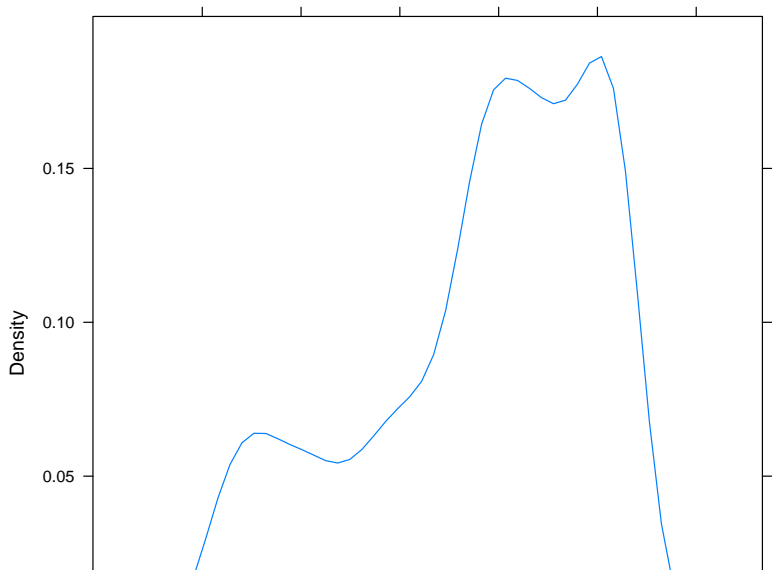
$$P(a < X < b) = \int_a^b f(x)dx$$

$$P(X < b) = \int_{-\infty}^b f(x)dx$$

$$P(X > a) = \int_a^{\infty} f(x)dx$$

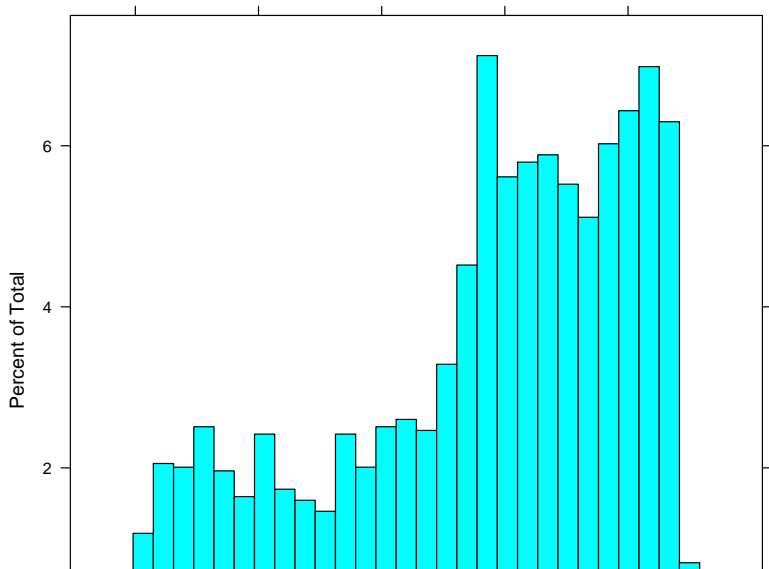
Función de Densidad de Probabilidad

Funcion de densidad de probabilidad



Histograma

Histograma



Media, varianza y desviación estándar

- ▶ La **media** de una variable aleatoria es el **centro de masas** de su función densidad de probabilidad:

$$\mu_X = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

- ▶ La **varianza** de una variable aleatoria es la **media del cuadrado de las desviaciones** respecto a la media:

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - \mu_X)^2 \cdot f(x) dx$$

- ▶ La **desviación estándar** es la raíz cuadrada de la varianza:

$$\sigma_X = \sqrt{\sigma_X^2}$$

Combinación lineal de variables aleatorias

- ▶ La **media de la suma** de varias variables aleatorias **independientes** es la suma de las medias:

$$\mu_{X_1+\dots+X_n} = \mu_{X_1} + \dots + \mu_{X_n}$$

- ▶ La **varianza de la suma o resta** de varias variables aleatorias **independientes** es la **suma** de las varianzas:

$$\sigma_{X_1\pm\dots\pm X_n}^2 = \sigma_{X_1}^2 + \dots + \sigma_{X_n}^2$$

Media y varianza de la media muestral

- ▶ Una **muestra de una población** es un conjunto de variables aleatorias independientes ($X_1 \dots X_n$).
- ▶ Si se toma una muestra de una población cuya media es μ y su varianza es σ^2 , entonces la media de la muestra es otra variable aleatoria (que es una suma de variables aleatorias)

$$\bar{X} = \frac{1}{n} \sum_n X_i$$

Media y varianza de la media muestral

- ▶ Por tanto, la **media de la media muestral** es la media de población:

$$\bar{X} = \frac{1}{n} \sum_n X_i = \mu$$

- ▶ La **varianza de la media muestral** es la suma de las varianzas:

$$\sigma_{\bar{X}}^2 = \sigma_{\frac{1}{n}X_1}^2 + \dots + \sigma_{\frac{1}{n}X_n}^2 = \frac{\sigma^2}{N}$$

Por tanto, una forma de **reducir la incertidumbre** es realizar la **medida en repetidas ocasiones**.

Mediana y cuartiles

- ▶ La **mediana** divide el conjunto de valores de la variable en **dos mitades** iguales (divide el area encerrada por la función densidad de probabilidad en dos partes iguales).
- ▶ Los **cuartiles** dividen este area en **cuatro** partes iguales.
- ▶ El area encerrada entre cada par de cuartiles es igual al 25\
- ▶ La **mediana** es el **segundo cuartil**.
- ▶ La **distancia intercuartil** (definida entre los cuartiles 1 y 3) es una **medida de la dispersión** de la variable.

① Bases de Datos

② Control de Calidad

Estadística

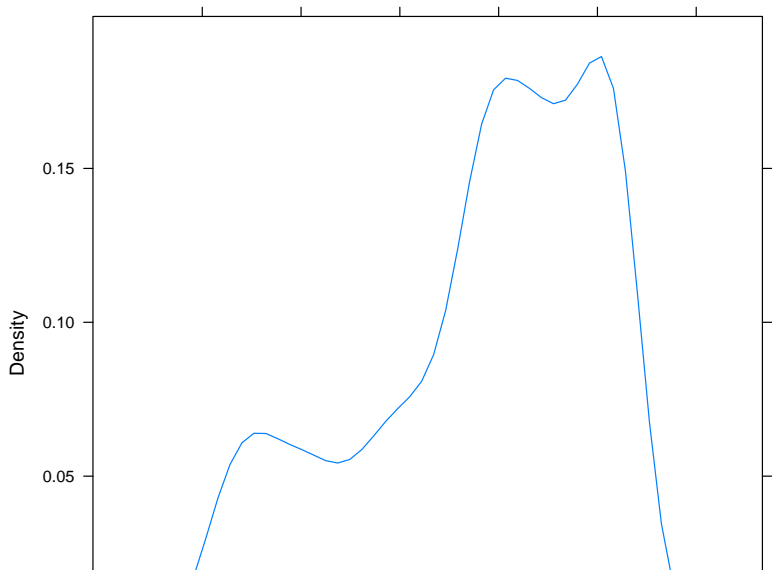
Gráficos

Control de Calidad de Medidas

Control de Calidad de Modelos

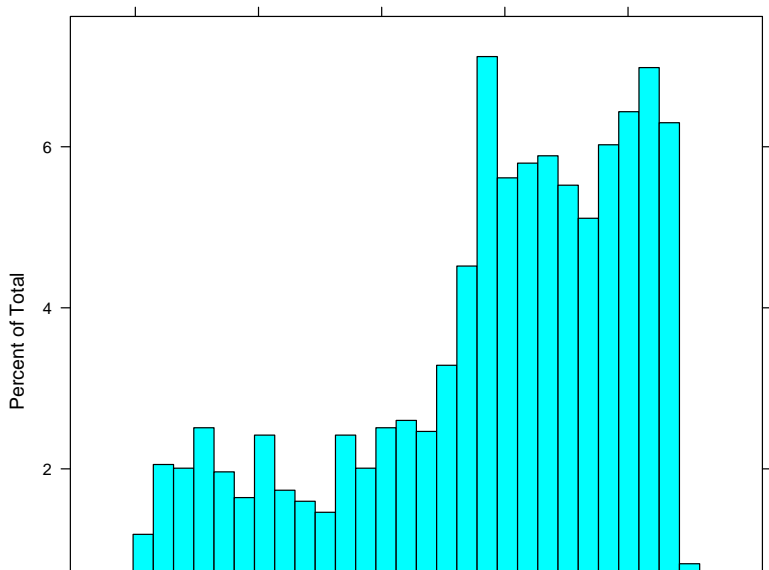
Función de Densidad de Probabilidad

Funcion de densidad de probabilidad



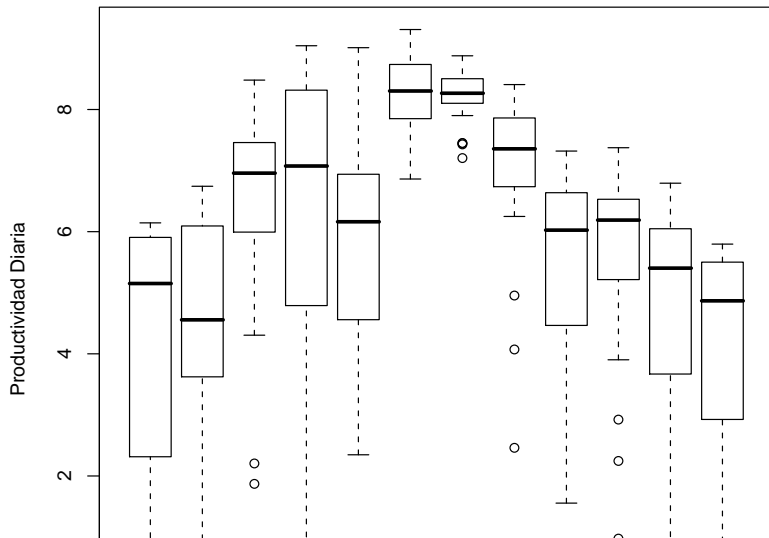
Histograma

Histograma



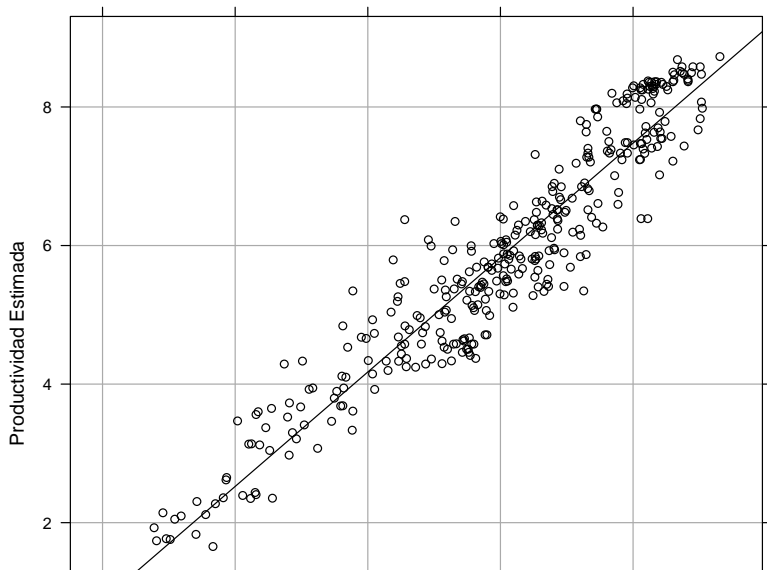
Gráficos boxplot

Variabilidad Mensual de la Productividad diaria

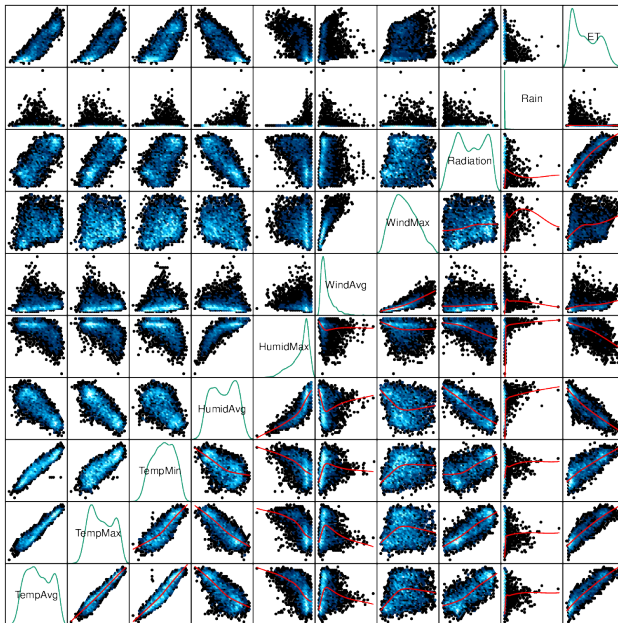


Gráficos de dispersión

Estimación de Productividad Diaria



Matrices de gráficos de dispersión



① Bases de Datos

② Control de Calidad

Estadística

Gráficos

Control de Calidad de Medidas

Control de Calidad de Modelos

Las medidas recogidas por estaciones meteorológicas se deben filtrar para eliminar datos erróneos.

- ▶ Límites Físicos
- ▶ Tests de persistencia
- ▶ Tests de rampas (irradiancia)
- ▶ Tests de envolvente (medida de varias componentes)
- ▶ Coherencia espacial
- ▶ Coherencia estadística

Irradiación Diaria

- ▶ La radiación global en el plano horizontal debe ser inferior a la extraterrestre ($K_t \leq 1$)

$$G_d(0) \leq B_{od}(0)$$

- ▶ El índice de claridad debe ser superior a 0.03

$$K_t = \frac{G_d(0)}{B_{od}(0)} \geq 0,03$$

- ▶ La radiación global en el plano horizontal debe ser inferior a la de un modelo de cielo claro

Irradiancia (intradiaria)

- ▶ El índice de claridad debe ser inferior a 1 cuando la altura solar es suficiente:

$$k_t < 1 \text{ si } \gamma_s > 2^\circ$$

- ▶ Límites inferiores para cielos cubiertos (baja transparencia atmosférica)

$$k_t \geq 10^{-4} \cdot (\gamma_s - 10^\circ) \text{ si } \gamma_s > 10^\circ$$

$$G \geq 0 \text{ si } \gamma_s \leq 10^\circ$$

Tests de persistencia

Variabilidad de irradiancia

- La media y la desviación estándar se calculan con todas las muestras de un día completo.

$$\frac{1}{8}\overline{k_t} \leq \sigma_{k_t} \leq 0,35$$

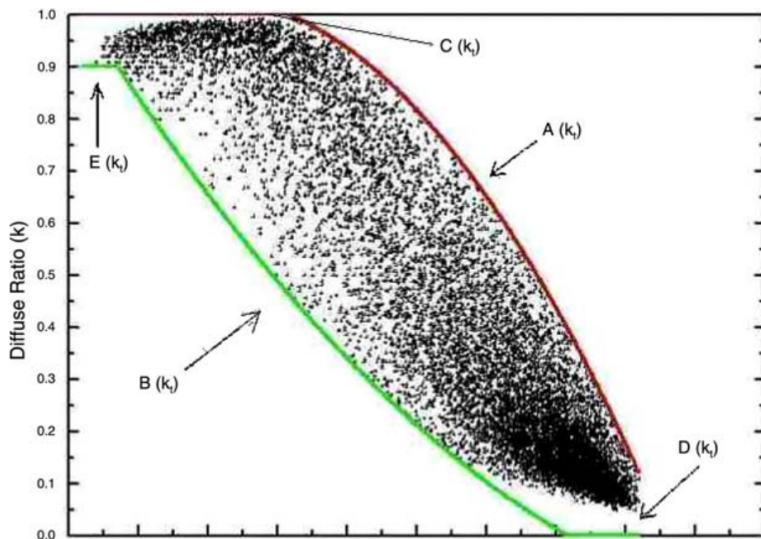
Tests de rampas

Límites a las variaciones de la irradiancia entre instantes sucesivos

$$|k_t(t) - k_t(t - 1)| < 0,75 \text{ si } \gamma_s(t) > 2^\circ$$

Tests de envolverte

- Sólo para estaciones con medida simultánea de global y directa/difusa.



Coherencia espacial

- ▶ Las medidas de una estación se pueden comparar con las recogidas por estaciones cercanas.
- ▶ Esta comprobación debe realizarse con **datos agregados** (diarios) (la variabilidad espacial intradiaria puede ser alta)
- ▶ Esta comprobación debe realizarse con estaciones que tienen **clima y geografía similar**.

Coherencia espacial

Pasos

- ▶ Estimamos la irradiación en el lugar, x_0 , con la interpolación espacial de las estaciones cercanas, x_i .
 - ▶ Los pesos w_i son una función inversa de la distancia (IDW).

$$\hat{G}_d(x_0) = \frac{\sum_{i=1}^N w_i G_d(x_i)}{\sum_{i=1}^N w_i}$$

- ▶ Comparamos la irradiación estimada, $\hat{G}_d(x_0)$, con la medida en la estación, $G_d(x_0)$.

$$\left| \hat{G}_d(x_0) - G_d(x_0) \right|$$

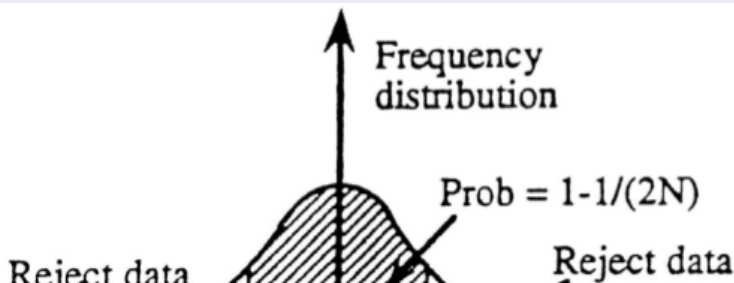
- ▶ La diferencia absoluta debe estar por debajo de un límite (p.ej. 50 %)

Coherencia estadística

Una medida puede ser etiquetada como *outlier* si es poco probable que pertenezca a la misma distribución que el conjunto.

Método de Chauvenet

Una medida es un *outlier* si la probabilidad de obtener su desviación respecto de la media es inferior al inverso de 2 veces el número de elementos en el conjunto.



Método de Chauvenet

- ▶ Sean $G_d(x_i)$ las medidas de radiación diaria del conjunto formado por N estaciones.

Método de Chauvenet

- ▶ Sean $G_d(x_i)$ las medidas de radiación diaria del conjunto formado por N estaciones.
- ▶ Se calcula la media, $\overline{G_d}$, la desviación estándar, σ_{G_d} .

Método de Chauvenet

- ▶ Sean $G_d(x_i)$ las medidas de radiación diaria del conjunto formado por N estaciones.
- ▶ Se calcula la media, $\overline{G_d}$, la desviación estándar, σ_{G_d} .
- ▶ Se calcula la distancia estadística de cada estación al conjunto:

$$d_i = \frac{G_d(x_i) - \overline{G_d}}{\sigma_{G_d}}$$

Método de Chauvenet

- ▶ Sean $G_d(x_i)$ las medidas de radiación diaria del conjunto formado por N estaciones.
- ▶ Se calcula la media, $\overline{G_d}$, la desviación estándar, σ_{G_d} .
- ▶ Se calcula la distancia estadística de cada estación al conjunto:

$$d_i = \frac{G_d(x_i) - \overline{G_d}}{\sigma_{G_d}}$$

- ▶ En una distribución gaussiana se calcula la distancia estadística equivalente a la probabilidad límite, $1/2N$, teniendo en cuenta las dos colas.
 - ▶ Por ejemplo, para un conjunto de 10 estaciones cada cola es $1/40 = 0,025$, el límite es $|d_{max}| = 1,96$.

Método de Chauvenet

- ▶ Sean $G_d(x_i)$ las medidas de radiación diaria del conjunto formado por N estaciones.
- ▶ Se calcula la media, $\overline{G_d}$, la desviación estándar, σ_{G_d} .
- ▶ Se calcula la distancia estadística de cada estación al conjunto:

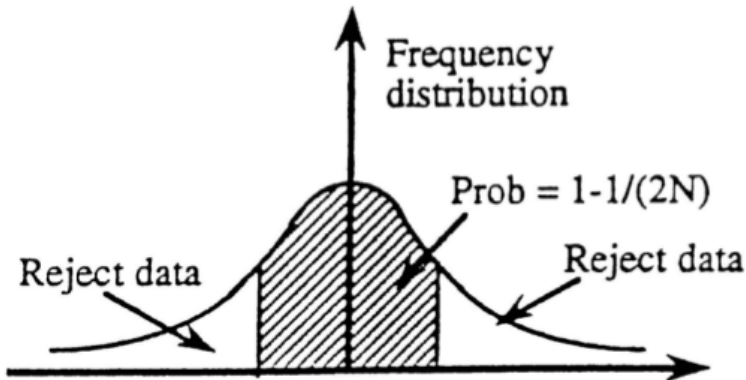
$$d_i = \frac{G_d(x_i) - \overline{G_d}}{\sigma_{G_d}}$$

- ▶ En una distribución gaussiana se calcula la distancia estadística equivalente a la probabilidad límite, $1/2N$, teniendo en cuenta las dos colas.
 - ▶ Por ejemplo, para un conjunto de 10 estaciones cada cola es $1/40 = 0,025$, el límite es $|d_{max}| = 1,96$.
- ▶ Aquellas observaciones que superan la distancia son marcadas como outliers.

Método de Chauvenet

$$d_i = \frac{G_d(x_i) - \bar{G}_d}{\sigma_{G_d}}$$

$$|d_i| > |d_{max}|$$



① Bases de Datos

② Control de Calidad

Estadística

Gráficos

Control de Calidad de Medidas

Control de Calidad de Modelos

Desviación entre modelo y observación

- Sea O el conjunto de observaciones (medidas) de una variable aleatoria.

$$\mathbf{O} = \{o_1 \dots o_n\}$$

- Sea M el conjunto de resultados de un modelo que aproxima el comportamiento de la variable medida.

$$\mathbf{M} = \{m_1 \dots m_n\}$$

- La desviación entre modelo y observación es:

$$\mathbf{D} = \mathbf{M} - \mathbf{O} = \{(m_1 - o_1) \dots (m_n - o_n)\} = \{d_1 \dots d_n\}$$

Estimadores frecuentes: MBD y RMSD

- Mean Bias Difference (MBD), diferencia media (indica si el modelo sobreestima o subestima):

$$MBE = \overline{\mathbf{D}} = \overline{\mathbf{M}} - \overline{\mathbf{O}} = \frac{1}{n} \sum_{i=1}^n (m_i - o_i)$$

Estimadores frecuentes: MBD y RMSD

- Mean Bias Difference (MBD), diferencia media (indica si el modelo sobreestima o subestima):

$$MBE = \overline{D} = \overline{M} - \overline{O} = \frac{1}{n} \sum_{i=1}^n (m_i - o_i)$$

- Root Mean Square Error (RMSD), diferencia cuadrático media:

$$RMSD = \left(\frac{1}{n} \sum_{i=1}^n d_i^2 \right)^{1/2} = \left(\frac{1}{n} \sum_{i=1}^n (m_i - o_i)^2 \right)^{1/2}$$

Estimadores frecuentes: MBE y RMSD

- Varianza de la diferencia (unbiased RMSD):

$$\sigma_{\mathbf{D}}^2 = \frac{1}{n} \sum_{i=1}^n (d_i - \overline{\mathbf{D}})^2$$

Estimadores frecuentes: MBE y RMSD

- Varianza de la diferencia (unbiased RMSD):

$$\sigma_{\mathbf{D}}^2 = \frac{1}{n} \sum_{i=1}^n (d_i - \overline{\mathbf{D}})^2$$

- El RMSD agrega información del promedio y la varianza de la diferencia:

$$RMSD^2 = \sigma_{\mathbf{D}}^2 + \overline{\mathbf{D}}^2$$

Otros estimadores: MAD

- Mean Absolute Deviation (MAD):

$$MAD = \frac{1}{n} \sum_{i=1}^n |d_i| = \frac{1}{n} \sum_{i=1}^n |m_i - o_i|$$

- El RMSD no es robusto (un error puntual puede distorsionar el estimador) y depende del número de muestras:

$$MAD \leq RMSD \leq n^{1/2} MAD$$

Otros estimadores: t y d

- ▶ t de Student (valores pequeños indican buen comportamiento del modelo)
 - ▶ Permite añadir intervalos de confianza a las diferencias entre modelo y observación

$$t = \left(\frac{(n-1)MBD^2}{RMSD^2 - MBD^2} \right)^{1/2}$$

Otros estimadores: t y d

- ▶ t de Student (valores pequeños indican buen comportamiento del modelo)
 - ▶ Permite añadir intervalos de confianza a las diferencias entre modelo y observación

$$t = \left(\frac{(n-1)MBD^2}{RMSD^2 - MBD^2} \right)^{1/2}$$

- ▶ d_1 : Índice de concordancia de Willmott.
 - ▶ Limitado entre 0 (ausencia de concordancia) y 1 (concordancia total).
 - ▶ Robusto frente a *outliers*.

$$d_1 = 1 - \frac{\sum_{i=1}^n |m_i - o_i|}{\sum_{i=1}^n (|m_i - \overline{\mathbf{O}}| + |o_i - \overline{\mathbf{O}}|)}$$

Correlación

El coeficiente de correlación entre dos conjuntos de datos es una medida numérica de la relación **lineal** entre los dos conjuntos (si la relación no es lineal, este coeficiente no sirve):

$$r = \frac{1}{n-1} \cdot \sum_{i=1}^n \left(\frac{o_i - \overline{\mathbf{O}}}{\sigma_{\mathbf{O}}} \right) \cdot \left(\frac{m_i - \overline{\mathbf{M}}}{\sigma_{\mathbf{M}}} \right)$$

Diagramas de Taylor

- ▶ Desarrollando σ_D^2 y teniendo en cuenta la definición de r :

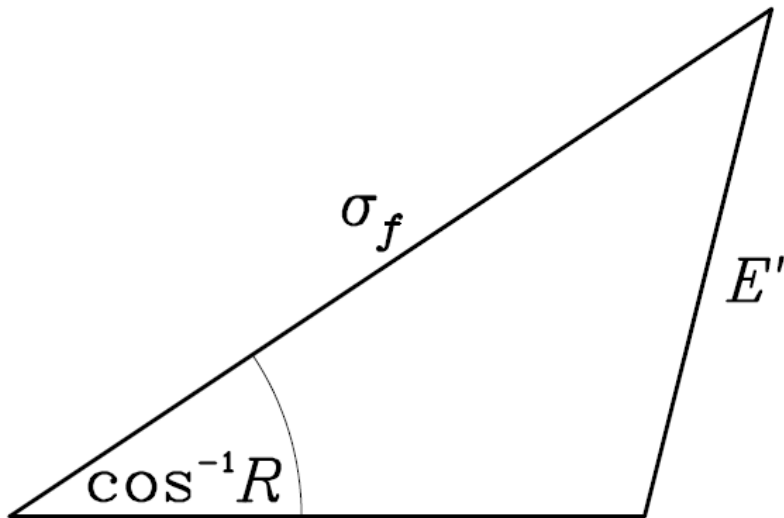
$$\sigma_D^2 = \sigma_O^2 + \sigma_M^2 - 2 \cdot \sigma_O \cdot \sigma_M \cdot r$$

- ▶ Esta relación es semejante a la ley de los cosenos (c, a, b son lados de un triángulo y ϕ es el ángulo opuesto al lado c):

$$c^2 = a^2 + b^2 - 2 \cdot a \cdot b \cos \phi$$

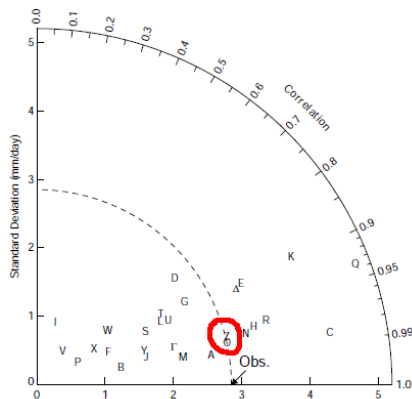
Diagramas de Taylor

$$\sigma_D^2 = \sigma_O^2 + \sigma_M^2 - 2 \cdot \sigma_O \cdot \sigma_M \cdot r$$



Diagramas de Taylor

- ▶ σ_D^2 : Distancia al origen
- ▶ σ_O^2 : Eje horizontal
- ▶ σ_M^2 : Eje vertical
- ▶ r : acimut



Target Diagram

- ▶ Emplea la relación entre $RMSD$, σ_D^2 , y \bar{D} , normalizadas con σ_O :

$$RMSD' = RMSD / \sigma_O$$

$$\sigma_D' = \sigma_D / \sigma_O$$

$$\bar{D}' = \bar{D} / \sigma_O$$

$$RMSD'^2 = \sigma_D'^2 + \bar{D}'^2$$

$$sign_\sigma = sign(\sigma_M - \sigma_O)$$

- ▶ Incorporan el signo de la diferencia entre desviaciones estándar de modelo y observación:

Target Diagram

- ▶ σ'_D (con signo): Eje horizontal
- ▶ \overline{D}' : Eje vertical
- ▶ $RMSD'^2$: Distancia al origen

