**Dataset**

Tendremos varios dataset de estaciones diferentes. Por ahora solo trabajaremos con uno. Lo que surja lo aplicamos al resto

**2-Prepro cesamiento**

Hay instancias que están a nivel de minutos, otros a nivel de cada 15 minutos, e incluso puede haber otras en otros periodos. Debemos uniformar y dejarlo todo por hora. En algunas variables para agregar por hora se debe promediar como en el caso de la temperatura, en otras se tendría que sumar como en el caso de leaf Wetness. Hay que tener mucho cuidado con esto, porque un error acá se trae abajo todo el resto del trabajo.

Además, debemos imputar los datos faltantes en algunas variables. Hay métodos de interpolación y otros métodos para imputar valores faltantes en series de tiempo.

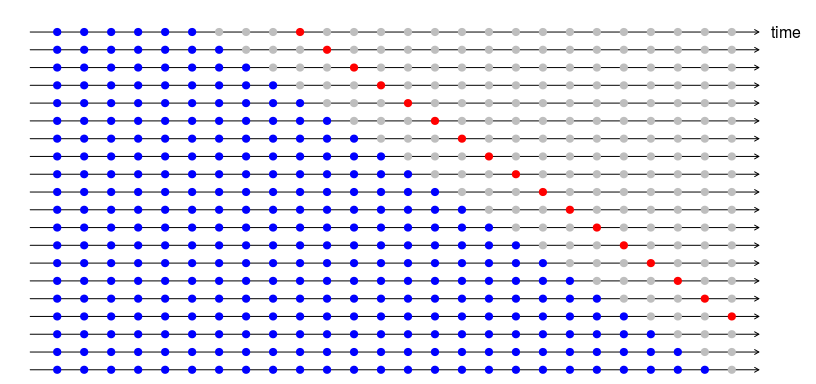
Cuando hay dos sensores tomar la media en las variables.

En general, revisar las variables para ver si hay datos que se vean extraños.

**3- Crear metodología de experimentación**

Nota/ Las series de tiempo se trabajarán con predicción recurrente de ser posible en vez de predicción directa. Recurrente significa que predice un periodo a la vez, y con base en la predicción anterior se hace la predicción futura. Esto es para que el modelo sea flexible con base en el periodo de tiempo que se necesite, a diferencia de la predicción directa. El método de validación cruzada puede ser prequential. Podemos usar el primer caso que se ve en el dibujo





4. Crear un modelo basado solo en series de tiempo para predicción que sea univariado (solo predice a futuro con base en el comportamiento de la serie). Podemos usar alguno de estadística y otro de Machine learning u aprendizaje profundo. Podría ser DeepAr, o algunos de los ganadores de la competencia M4

5. Luego podemos ver si creamos un algoritmo con base en Deep learning que considere todas las variables simultáneamente, como la temperatura del aire , la humedad relativa, y otras. Y las prediga a futuro. Siempre teniendo en cuenta que nuestra variable mas importante es leaf wetness