Supplement of Geosci. Model Dev., 17, 6745–6760, 2024
https://doi.org/10.5194/gmd-17-6745-2024-supplement

Geoscientific Model Development
Open Access
EGU

Supplement of

The Measurement Error Proxy System Model: MEPSM v0.2

Matt J. Fischer

Correspondence to: Matt J. Fischer (mjf@ansto.gov.au)

RI6prep

Matt Fischer

March 21, 2023

[1]:
```
import Cairo
using CSV, DataFrames, Dates, Statistics, MAT
using Gadfly
```

[2]:
```
ENV["COLUMNS"] = 300
```

[2]:
```
300
```

[3]:
```
a = matread(expanduser("~/Ddrv/Data/PAGES/CoralHydro2k1_0_0.mat"))["TS"];
```

[4]:
```
# "paleoData_hasResolution_hasMeanValue",  "paleoData_samplingResolution", "yearUnits", "datasetId",
 ↪"paleoData_notes"

col_str=[ "dataSetName", "archiveType", "geo_meanElev", "geo_meanLat", "geo_meanLon", "paleoData_coralHydro2kGroup",
 ↪"hasResolution_nominal",
    "paleoData_variableName", "paleoData_units", "paleoData_analyticalError"
]
```

[4]:
```
10-element Vector{String}:
 "dataSetName"
 "archiveType"
 "geo_meanElev"
 "geo_meanLat"
 "geo_meanLon"
 "paleoData_coralHydro2kGroup"
 "hasResolution_nominal"
 "paleoData_variableName"
 "paleoData_units"
 "paleoData_analyticalError"
```

[5]:
```
Ds0 = DataFrame(Dict(k⇒vec(v) for (k,v) in a if in(k, col_str)))
Ds0 = ifelse.(isempty.(Ds0), missing, Ds0)
Dsite = Ds0[Ds0.paleoData_variableName.≠"year",:]
Dsite.lon2 = Dsite.geo_meanLon + (Dsite.geo_meanLon.<0.)*360
first(Dsite, 6)
```

[5]:

	archiveType	dataSetName	geo_meanElev	geo_meanLat	geo_meanLon	hasResolution_nominal	paleoData_analyticalError	paleoData_coralHydro2kGroup	paleoData_units	paleoData_variableName
	String	String	Float64?	Float64	Float64	String?	Any	Any	String	String
1	coral	AB08MEN01	-6.0	-0.13	98.52	monthly	0.07	4.0	permil	d18O
2	coral	AB15BHB01	-0.5	-6.53	105.63	monthly	0.049	4.0	permil	d18O
3	coral	AB20MEN01	-0.7	-3.18	100.517	monthly	0.05	4.0	permil	d18O
4	coral	AB20MEN02	missing	-2.37	99.745	monthly	0.05	4.0	permil	d18O
5	coral	AB20MEN03	missing	-3.126	100.309	monthly	0.05	4.0	permil	d18O
6	coral	AB20MEN04	missing	-2.752	99.995	monthly	0.05	4.0	permil	d18O

[6]:
```
# extrema(Dsite.lon2)
```

[7]:
```
i1 = Dsite.dataSetName .== "OS14RIP01"
Dri = Dsite[i1, :]
Dri.daterange = [[Date(1950,2), Date(2008, 3)]]
Dri
```

[7]:

	archiveType	dataSetName	geo_meanElev	geo_meanLat	geo_meanLon	hasResolution_nominal	paleoData_analyticalError	paleoData_coralHydro2kGroup	paleoData_units	paleoData_variableName
	String	String	Float64?	Float64	Float64	String?	Any	Any	String	String
1	coral	OS14RIP01	-2.0	7.2708	134.384	monthly	0.05	4.0	permil	d18O

[8]:
```
module aa

using DataFrames, Dates, Statistics
import Measurements as Mm
import Measurements: ±


function readTS(s::String, a)
    i = a["dataSetName"].==s
    z = a["paleoData_values"][i]
    derr = Dict(k⇒v for (k, v) in zip(a["paleoData_variableName"][i], a["paleoData_analyticalError"][i]))
```

```
    @show derr
    time = z[2][:]
    date1 = datef(time)
    D = DataFrame(id=s, date=date1, t=time, y=z[1][:] .± derr["d18O"])
    return D
#    return D[D.t.>1950,:]
end


function datef(time::AbstractVector)
    y = floor.(Int, time)
    m =  (time-y) .÷ median(diff(time)) .+ 1
    return Date.(y,m)
end

end
```

[8]:
```
Main.aa
```

[9]:
```
DOc = aa.readTS("OS14RIP01", a)
@show size(DOc)
first(DOc, 6)
```

```
derr = Dict{String, Any}("year" => "NA", "d18O" => 0.05)
size(DOc) = (1309, 4)
```

[9]:

	id	date	t	y
	String	Date	Float64	Measurem…
1	OS14RIP01	1899-03-01	1899.17	-5.546 ± 0.05
2	OS14RIP01	1899-04-01	1899.25	-5.729 ± 0.05
3	OS14RIP01	1899-05-01	1899.33	-5.892 ± 0.05
4	OS14RIP01	1899-06-01	1899.42	-5.792 ± 0.05
5	OS14RIP01	1899-07-01	1899.5	-5.728 ± 0.05
6	OS14RIP01	1899-08-01	1899.58	-5.676 ± 0.05

[10]:
```
Dri
```

[10]:

	archiveType	dataSetName	geo_meanElev	geo_meanLat	geo_meanLon	hasResolution_nominal	paleoData_analyticalError	paleoData_coralHydro2kGroup	paleoData_units	paleoData_variableName
	String	String	Float64?	Float64	Float64	String?	Any	Any	String	String
1	coral	OS14RIP01	-2.0	7.2708	134.384	monthly	0.05	4.0	permil	d18O

[11]:
```
module t

using DataFrames, Dates, Statistics
using Rasters
using Rasters: Between
import Rasters.DimensionalData as DD
import Measurements as Mm
import Measurements: ±


function oceandata(fname::String, varname::String; Dsite=DataFrame(), k=Date[], f1="")
        fname = f1*fname
    f2 = splitdir(fname)[2]
    valname = Symbol(split(f2, '.')[1])
    ds0 = Raster(fname, key=varname)
    ds = set(ds0, Ti⇒DD.Points)
    if isempty(k)
            k = [[Date(year(x[1]), month(x[1]), 1), Date(year(x[2]), month(x[2]), 28)] for x in Dsite.daterange]
    end
    lonmax = maximum(dims(ds, X))
    sitelon = lonmax>180 ? Dsite.lon2 : Dsite.geo_meanLon
    dsst = [DataFrame(id=id, date=t1[1]:Month(1):t1[2], t=timef(t1),
            sst=vec(read(view(ds, X(Near(x)), Y(Near(y)), Ti(Between(t2...)))))))
            for (id, x, y, t1, t2) in zip(Dsite.dataSetName, sitelon, Dsite.geo_meanLat, Dsite.daterange, k)]
    D2 = reduce(vcat, dsst)
    if any(skipmissing(D2.sst.<0.0))
        allowmissing!(D2, :sst)
        D2[findall(skipmissing(D2.sst.<0)), :sst] .= missing
    end
    if any(skipmissing(D2.sst.==0.0))
        D3 = combine(groupby(D2, :id), [:t, :sst]⇒seasonalize⇒:y)
        D2[!,:sst] = D3.y
    end
    rename!(D2, :sst⇒valname)
    return D2
end
```

```
function timef(tv::Vector{Date})
    date1 = tv[1]:Month(1):tv[2]
    return year.(date1) + month.(date1)./12 .- 1/24
end


function createDF1(Dsn::DataFrame; Dsite=DataFrame(), f1="")
    Dz = oceandata.(Dsn.file, Dsn.var, Dsite=Dsite, f1=f1)
    D1 = Dz[1]
    vname = propertynames(D1)[end]
    goodi = .!ismissing.(Dz[2][:,end])
    D2 = copy(D1[goodi,:])
    D2[!, vname] = D1[goodi, end] .± Dz[2][goodi, end]
    vnamea = Symbol(vname, "a")
    return D2
end


function seb(X::Vector, mse::Float64)
    μx = mean(X)
    svb = mse / sum(abs2, X.-μx)
    return sqrt.(svb)
end



function detrend(t1::AbstractVector, Yo::AbstractVector)
    X = [ones(length(t1)) t1.-mean(t1)]    # no error in X
    n, p = size(X)
    Xt = permutedims(X)
    Y = Mm.value.(Yo)
    b = Xt*X \ Xt*Y
    mse = sum(abs2, Y - X*b)/(n-p)
    b2 = b[2] ± seb(X[:,2], mse)
    @show b2
    Yh = X[:,2]*b2
     return Yo-Yh
end


end
```

[11]:
```
Main.t
```

[12]:
```
f4 = expanduser("~/Ddrv/Data/Oceans/Salinity/")
Dsssn = DataFrame(file=["HadEN4.2.1g10_190001-201012.nc", "HadEN4.2.1g10_sss_uct.nc"], var=["sss", "salt_err_std"])
Dsss = t.createDF1(Dsssn, Dsite=Dri, f1=f4)
first(Dsss, 4)
```

[12]:

	id	date	t	HadEN4
	String	Date	Float64	Measurem…
1	OS14RIP01	1950-02-01	1950.12	34.03 ± 0.17
2	OS14RIP01	1950-03-01	1950.21	34.01 ± 0.17
3	OS14RIP01	1950-04-01	1950.29	33.98 ± 0.17
4	OS14RIP01	1950-05-01	1950.38	33.95 ± 0.17

[13]:
```
f1 = expanduser("~/Ddrv/Data/SST/")
Dsstn = DataFrame(file=["ersstv5.nc", "ersstev5.nc"], var=["sst","ut"])

Dsst = t.createDF1(Dsstn, Dsite=Dri, f1=f1)
first(Dsst, 4)
```

[13]:

	id	date	t	ersstv5
	String	Date	Float64	Measurem…
1	OS14RIP01	1950-02-01	1950.12	28.15 ± 0.31
2	OS14RIP01	1950-03-01	1950.21	28.35 ± 0.27
3	OS14RIP01	1950-04-01	1950.29	28.92 ± 0.29
4	OS14RIP01	1950-05-01	1950.38	29.1 ± 0.27

[14]:
```
Docn = innerjoin(Dsst, Dsss[:, Not(:t)], DOc[:, Not(:t)], on=[:id, :date])
@show size(Docn)
first(Docn, 4)
```

[14]:
```
size(Docn) = (695, 6)
```

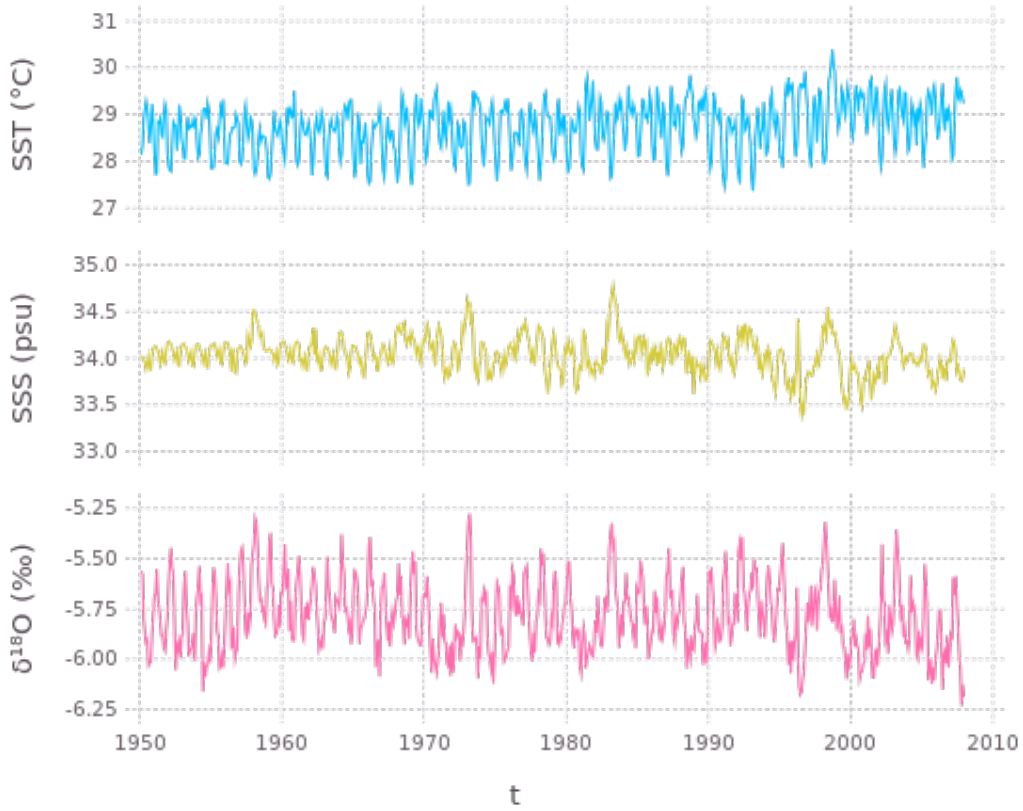
	id	date	t	ersstv5	HadEN4	y
	String	Date	Float64	Measurem…	Measurem…	Measurem…
1	OS14RIP01	1950-02-01	1950.12	28.15 ± 0.31	34.03 ± 0.17	-5.562 ± 0.05
2	OS14RIP01	1950-03-01	1950.21	28.35 ± 0.27	34.01 ± 0.17	-5.579 ± 0.05
3	OS14RIP01	1950-04-01	1950.29	28.92 ± 0.29	33.98 ± 0.17	-5.804 ± 0.05
4	OS14RIP01	1950-05-01	1950.38	29.1 ± 0.27	33.95 ± 0.17	-5.871 ± 0.05

[ ]:

[15]:
```
Dv = select(Docn, :t, [4:6;].⇒ByRow(t.Mm.value), renamecols=false)
Dvl = stack(Dv, Not(:t));
```

[16]:
```
yglabs = Dict("ersstv5"⇒"SST (°C)", "HadEN4"⇒"SSS (psu)", "y"⇒"δ¹⁸O (‰)")

p1 = plot(Dvl, x=:t, y=:value, ygroup=:variable, color=:variable,
    Geom.subplot_grid(Geom.line, free_y_axis=true),
    Scale.ygroup(labels=i->yglabs[i]),
    Guide.ylabel(nothing),
    Theme(key_position=:none)
)
draw(PNG(6inch, 5inch), p1)
```



[16]: false

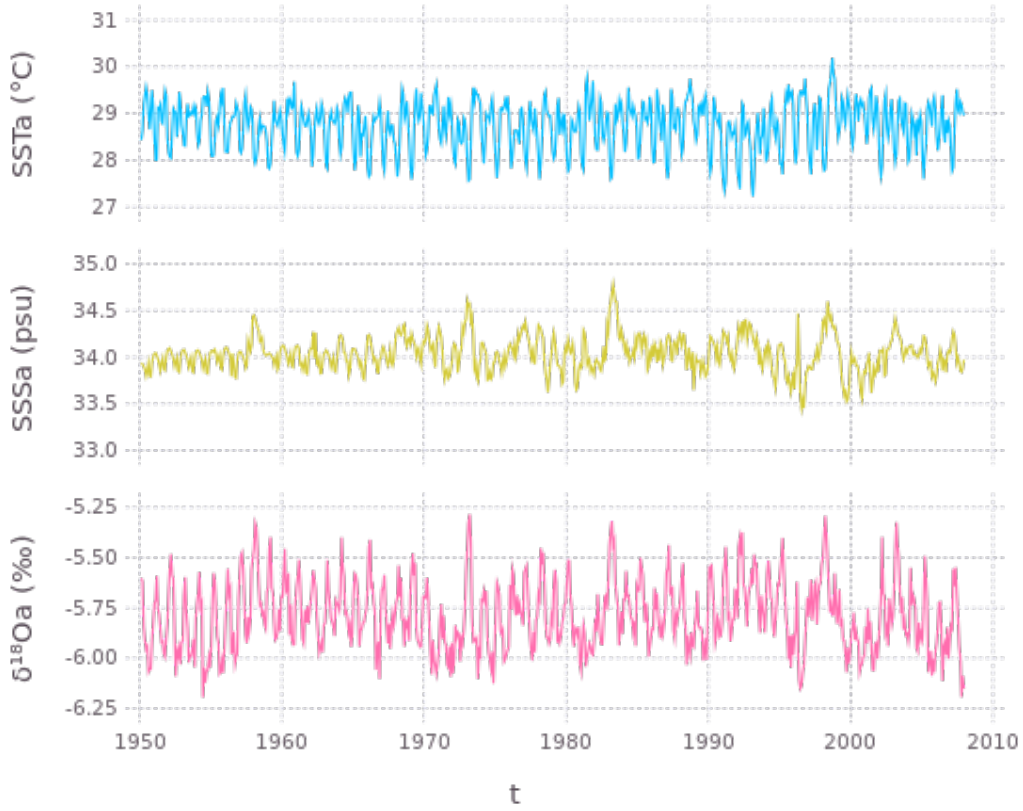
[ ]:

[17]:
```
Docna = select(Docn, :t, [4:6;].⇒(y->t.detrend(Docn.t, y))⇒x->string(x,"a"))
 Dv2 = select(Docna, :t, [2:4;].⇒ByRow(t.Mm.value), renamecols=false)
Dv2l = stack(Dv2, Not(:t)) ;
```

```
b2 = 0.0098 ± 0.0012
b2 = -0.00295 ± 0.00042
b2 = -0.00135 ± 0.00039
```

[18]:
```
yglabs = Dict("ersstv5a"=>"SSTa (°C)", "HadEN4a"=>"SSSa (psu)", "ya"=>"δ¹⁸Oa (‰)")

p2 = plot(Dv2l, x=:t, y=:value, ygroup=:variable, color=:variable,
    Geom.subplot_grid(Geom.line, free_y_axis=true),
    Scale.ygroup(labels=i->yglabs[i]),
    Guide.ylabel(nothing),
    Theme(key_position=:none)
)
draw(PNG(6inch, 5inch), p2)
```



[18]:
```
false
```

[19]:
```
# CSV.write("./Data/Docna_1950_v100.csv", Docna)
# CSV.write("./Data/Palaui.csv", Docna)
```

[ ]:

[ ]:

[ ]:

[ ]:

[ ]: