DDAHS 2026 summer school Workshop 4 Researching Hydrological Data-20260715_133518-Meeting Recording

15 July 2026, 12:35pm

2h 48m 42s

◉ **Gianni Vesuviano** started transcription

Tom Keel 0:10

Alright, so the first thing I'll try is I'll share my screen.
This is not for some issues.
Should have done this before.
Can you see my screen?

Matt Dalle Piagge 0:30

Yep.
Looks good. I like the background.

Tom Keel 0:34

It's just something random I found last year.

Matt Dalle Piagge 0:37

Looks like the Lake District. Sorry.

Tom Keel 0:39

It could be.
Yeah, I wanted to try to like visualise rain, rainy, rainy valley. This is just something that kind of came up. Okay, so hi everyone. Welcome to workshop 4. I'm Tom Keel. So I've been in some of the other sessions. I've introduced myself. I'm A hydrological data analyst.
basically just do data science and a bit of software development as well. Mainly looking at so far rainfall data, something I've always been interested in since I was a kid, basically. I am going to go through a workshop looking at how we can sort of do the practical

like set up practical research workflows. So where the workshop today may be a bit advanced for those of you that aren't familiar with Python, it's more so the key takeaway from today will be, well, how do you actually set up like a PhD research level workflow?
to start off with some data, clean it, set up and run things. As Matt brilliantly went through before, like there's loads of different types of data that we have for hydrological sciences. And like a lot of it is so big that we need to have it stored somewhere in the cloud so we can just get a small subset of it and then work with that and then build up our analysis. But that step
taking a single CSV of one catchment or a single catchment of data, running some analysis and scaling that up to bigger and bigger study areas is something that in and of itself is a real skill and something that can be quite awkward to do. So that's, we'll talk about that scaling up. So we're already a little bit over. I don't know why I didn't account for
a longer time to talk, so this gives me not much time to do this overview. So I'll start off with an overview lecture, and then we'll just do a very brief setup and start the practical. Then I'll let you loose to do that practical until about half two. We'll have two 10-minute breaks and then go through a feedback. Feedback where I'll run through what you've just done in the practical,
start the next bit and then let you work again for half an hour and then come back for some more feedback. And then right at the end, I mean, essentially the whole thing is quite independent. But this is very much where you can design your own or give you some options for designing your own small research task. That's just taking some data
and starting to set up a research flow with that data. So there's three different tasks with a traffic light system of more easy to more difficult. And with all of these, because we only have half an hour and knowing these things will probably have even less than half an hour, we really just want you to get you thinking about how you would approach using this sort of data.
So I'll start off with a question to introduce the workshop, which is like, how do we know how much it rains? And are there uncertainties with that? So I guess like, I'll just wait for one person to respond, ideally verbally that chat is fine, or just some sort of ideas of like, how do we know how much it rains? Like, so what sort of?
Questions or answers to for that question, a very simple question.

Jones, Robyn (MECP) 4:00
We can measure it with rain gauges.

Tom Keel 4:02
Brilliant. Fantastic. Yeah. So that's what I was hoping would be the only answer, not something too complicated or awkward. But basically, yeah, we have ways of measuring rainfall. But then again, there's uncertainty. It's like, well, yes, we can measure rainfall with, as you say, rain gauges.
But that like is in one specific point, it has a specific resolution. So that data may be collected at the end of at the end of a day by someone or start of a day could be collected. There's even gauges in the UK that someone will go check on them every every month. And then you have things that are automated.
In each case, we have.
uncertainties introduced to our picture of rainfall. I mean, if we just took this example here, how much it is raining in, I don't know, on this person here versus this person here, there is uncertainty spatially because of the way that rainfall works. So we really, in the UK, when we build up
data products, we really have to try to think about these sorts of things. That's what the practical is going to be about as well. So what makes the data uncertain? So one of the first sources of uncertainty is the fact that we have loads of different sources. So there's rain gauges, which has already been pointed out, which is essentially just buckets. We also have in the UK have radar.
And then we have data that tends to be used less for sort of small regional scale analysis, but we do have satellite data which has some very, very wobbly estimate of rainfall. But the fact that we have different sources of data creates a real problem because in some cases, rain gauges is going to give you a point,
radar data gives you a field. So that creates uncertainty when we're trying to blend them together. There's also, so that the UK is obviously like a rich developed nation and we have quite high density of observations that go back a long time. We also have a good, or have had a good network of sort of people to maintain those.
But there are notable exceptions. For example, in the north of Ireland and Scotland, there is these areas where you have a less dense network. So we're less certain about what the rainfall is there. And similarly, I believe there's only maybe 2 radar spots in Scotland that capture the rainfall. So we have lots of uncertainty because of the

density. And then finally, something I've worked on a lot at UKCEH is like,
OK, fair enough. We have this data. Maybe we don't care about how it's collected. It is what it is. But there is also a massive uncertainty introduced by the way in which we remove data from that. What sort of ways do we summarise spatial and temporal information? So this is very much another uncertainty. So all of these conflate to create very difficult pictures of
Well, how much it rains. So what data is available in the UK? So we'll be working today with the UK Rain Gauge Network. This is just a map of, I think this is from 2015, so it's probably a bit out of date, but it would look similar if we took the profile now. But essentially, this is collected by the Scottish Environment Protection Agency, the Met Office,
Natural Resources Wales and the Environment Agency. So as many of you know, the UK is made-up of multiple different countries. They all have different operating agencies which collect data in separate ways. And essentially, for the UK, we rely on, or hydrological modellers would rely on this thing of taking
This quite and sparsely populated some or like.
non-uniformly distributed rain gauges and we put these onto a grid. And so we have products like this. So this is directly from the CEH-GEAR, which Matt was talking about in the workshop before. And it's essentially just it gives you that smooth rainfall field from point data. And yeah, there's two different types in the UK.
If you don't know already and you ever wanted to work with UK data, there's the HadUK-Grid produced for the Met Office that is updated yearly, it's daily and monthly. And then we have the CEH-GEAR, which I'm working on a replacement for hourly, so 15 minute data, but currently this only goes up to 2021, I think.
But yeah, basically use this one if you're looking at it for now. So yeah, but this introduces this fourth type of uncertainty, which is brilliant. So like, how do we know if we were to, if we lived in, or where are we right now? We're in Oxfordshire, Wallingford, or I am. We looked at that grid cell versus the nearest gauge, what would the differences be?
that's going to completely tell a completely different story. Okay, so today's workshop is we're going to zoom in on one region, that is the Upper Severn, and we're just going to ask a simple question, how much does it rain in the Upper Severn? So the Upper Severn is based between Wales and England. It's the longest river in the UK, I think. I just looked it up.
before they started. And essentially, we're going to look at a relatively arbitrary sub-

catchment area of that, and that's just the very upper bits in the Cambrian Mountains, so to the west of Shrewsbury here.
What might rate rainfall estimates uncertainty? So for one thing, yes, in the Cambrian Mountains, so it's topographically diverse. So again, we might have a rain gauge in the bottom of the valley, but the difference in rain at the top of the top of the hill, for example, is not going to be negligible. It's going to be important. It rains, tends to rain, or the weather's more variable.
at the top of these sorts of mountains or high topographical features. And then equally we have, so this is the same catchment outline shapefile. And we have quite a sparse network, so maybe our estimates, our gridded estimates aren't very certain there. So this is the data we work with over the Upper Severn. And we're going to be using Polars, which
I think Sebastian mentioned before, I'm glad that someone's heard of this before, but essentially it's just like Pandas, but it's just a much, much quicker version. It has a slightly different syntax of Pandas. So it reads in tabular data. That's essentially what it's doing. It reads in things like CSV, Parquet, GeoPandas, which we've been introduced to before, and Xarray, which we've been
introduced to before. So just the difference here is GeoPandas is for like geospatial, like geo coordinates data, some things with XY, also lat long and stuff like that. Xarray reads in three-dimensional data sets or four-dimensional data sets with time as well.
About this practical, so it's slightly different syntax. So you'll see throughout these things, this little smiley, confused man person, and you see that there's tasks and these are things that you have to do during the practical in this independent work, which I'm about to unleash you all to do. And there's also these optional tasks for the people that really want to
push themselves and learn more about, well, mainly Polars, I guess, but there's some optional tasks that really start to get you thinking, but these aren't essential. And just the reason I've had this as a slide is just to say you won't be able to just run this code. You do have to delete these question marks and...
Put in an answer, basically.
Very quickly as well, and I didn't realise until we started doing workshop 3, I didn't realise that Gemeni on Google Colabs sort of does your code for you. And I would be very, I just want to put a slight warning at the top. And just for all sort of, I mean, we all know this already, we live in this world, but

just for solving any task, if you're really trying to learn, I feel like the learning is like when you go up a mountain and an LLM like Gemini or the autocomplete is going to just jump you straight to the top of that mountain. You might not like pick up the sort of essential bits of learning to get you there, basically.
I'm not saying that it's not useful and research it very much is. I think you just have to be, with this learning, careful, really, really careful with yourself. And it doesn't matter if you ask other human beings in the chat here, silly, silly questions or basic questions. I think that's really how we can get a bit further.
And especially for research, very, very often you have to do something that is like going to take you down a very specific path that's unique to your case study and your region, your data. And a LLM might give you a answer that's a bit more generic and it doesn't actually apply to your specific research case. Anyway, that's just
I just added that because... But yes, okay, so then I'll give you some advice for doing data-driven research. These are the things that I want you to ideally take away, if nothing else, if not in the code. So essentially, when you're doing data-driven research, first thing you need to do is look at that data. So when we were answering, Robin's question earlier said like just have a look at the data. I think that's very much the very first thing.
to do if you've got a question, oh, I've got an error. So just see what the data looks like. And most often not, yeah, it's filled with NaNs or there's a clear and obvious error. Keep your code organized. This is something that like even I have been coding for many, many years, but like it's something that I have to keep coming back to. It's so easy to make shortcuts and not keep.
organised, but if you don't keep things organised, you're less likely to want to open your research and you'll get less far with your research. Essentially, it's as simple as that, like you just get scared of going back to something you were working on before. Working with subsets, which is something we'll do here and it's something that's essential, and Matt introduced it with
the object store of like if you try to load all of that data and that's terabytes and terabytes of data, it's not valid as well as when you're doing things with sort of like when you're saying going into using like machine learning algorithms as we will with Molly Asher's workshop tomorrow, it's a really good idea to just do a small chunk.
to check if a model works and then apply it to, yeah, loads of catchments. So working with subsets is a good idea. Write reusable functions. Matt kindly added some of his own for the last one. So just things that you can load in from elsewhere to use in

multiple ways. And then the one that I think is just the most important principles, a very software development principle, but is this idea of everything you do with code is should be following a single responsibility principle. That is, I like if someone to link the Wikipedia to that would be fantastic as well. But that is essentially everything you do should have one purpose. So if you're writing one variable, that should store one thing. If you're writing one function, that should do one thing. If you're writing one notebook, that should do one thing. So in this case, it kind of works because the notebook is 1 workshop. I guess it's not all the workshops put into one massive notebook. But this sort of principle scales a lot of things to the way we organize.
And develop code, but yeah, single responsibility principle is something I'd really. Suggest work for research.
Okay, so now enough talking. That's good. We've gone over a little bit. So we will start our independent work. So if someone could share the link to the workshop, I will give you sort of maybe 10 minutes while we put you into breakout rooms, and then we'll come round and we'll come back after 30 minutes. So
Yeah, please let us know if you can't get access to the materials. But yeah, there's some reading at the top, but then it's pretty much you'll just be straight into the coding after that.
So, 30-minute timer on this one.

Yelesen, Sebastian 15:40
You did it through an embedded YouTube video?

Tom Keel 15:43
Yeah, that's yeah, I've got a timer on my phone, yeah.

Yelesen, Sebastian 15:45
It’s definitely...
It's quite an interesting way to do it.

Gianni Vesuviano 15:50
Tom, with the breakout rooms, five people who are not using organisation managed Teams accounts cannot be put into any breakout room.

Tom Keel 15:53
Yeah, should they stay in here then? Bit scary, but...

Gianni Vesuviano 16:00
So.
Okay, and then...
And then should I move three groups of four?

Tom Keel 16:09
Yes, that'd be fantastic. Yeah.

Tom Keel 16:13
Obviously, is it can they join the main group if they'd like as well?

Gianni Vesuviano 16:20
Yhey should be able to leave the breakout rooms as whenever they want.

Tom Keel 16:25
Fantastic, yeah, cool.

Gianni Vesuviano 16:29
OK, they're all assigned, so I'm gonna open the rooms now.

Amulya Chevuturi 16:35
Matt, I call room 2? Are you room 1?
Okay.

Matt Dalle Piagge 16:42
What, whichever?

Tom Keel 16:42
Give everyone like 5 minutes or so just to set up and just, yeah, familiarize.

Amulya Chevuturi 16:48
Okay, yeah, we can stay.

Amulya Chevuturi 17:17
We might have to go just in case they have like...
Questions about which section, what section?

Tom Keel 17:25
Oh, sure, yeah. I mean, well, I mean, hopefully they'd ask each other. But maybe there's a more fundamental issue where they can't get it open or something.

Amulya Chevuturi 17:35
Gianni, I can't join any rooms, so you'll have to set up that. Can I join rooms? I can't see any rooms.
Okay, first you finish setting up the room, sorry.

Gianni Vesuviano 17:47
I, I have set up the rooms.

Amulya Chevuturi 17:49
Wait, there are a couple of people still in the main room.

Gianni Vesuviano 17:52
Yeah, I said five people can't get into the rooms because they don't use organisational Teams accounts.

Amulya Chevuturi 17:58
Okay, so they are staying in main. Okay, sorry, I got confused.

Gianni Vesuviano 18:03
Yeah, but you're an organizer, so you should be able to move yourself.

Amulya Chevuturi 18:07

Oh no, I can see the rooms. Yeah, sorry. Yes. I'm calling to room 2.
Bye.

Tom Keel 18:15
Right, for people staying, was everything OK? To like, can you set it up everything OK, or just message in the chat and mic unmute?

Gianni Vesuviano 18:32
Matt or Kit, are you going to move into the...

Tom Keel 18:34
Fantastic.

Gianni Vesuviano 18:37
One of the other breakout rooms.

Matt Dalle Piagge 18:40
Yes.
But I'm going to be a Teams noob and say that I can't work out how to do that.

Kit Macleod 18:48
in the top bar in the rooms, click on rooms and then it should say, you know, like in the right hand side panel, the ones you can join.

Matt Dalle Piagge 18:52
Hi.
Can't the problem is, I can't see rooms in the top panel.

Kit Macleod 18:57
Okay.

Tom Keel 18:58
Neither can I, actually.

Kit Macleod 18:59
Okay, for me it's between notes and polls. Okay.

Matt Dalle Piagge 19:03
No, not for me.

Tom Keel 19:04
Oh no, this might be a Linux thing. Damn it. Yeah.

Kit Macleod 19:04
Okay.

Matt Dalle Piagge 19:07
It probably is, isn't it Tom?

Kit Macleod 19:07
Okay.
Okay.

Matt Dalle Piagge 19:11
Gianni, are you able to add me to a room or push me into a room or whatever it is you have to do?

Gianni Vesuviano 19:17
Yeah, give me a moment. I'll have to take you off as an organizer.

Matt Dalle Piagge 19:21
Okay, sure.

Gianni Vesuviano 19:22
Right.
Okay, I can assign you. I'm going to put you into room one.

Matt Dalle Piagge 19:54
Okay.

Kit Macleod 19:54
Where did Amulya go?

Matt Dalle Piagge 19:57
Room 2, I think.

Gianni Vesuviano 19:57
So you should go into room 3 because Amulya is in room 2.

Kit Macleod 20:01
Yeah.

Tom Keel 20:06
Matt... oh, yeah, he's got it's fine.

Matt Dalle Piagge 20:20
Uh, no, I can I can see it in the top left of my screen, it's...
On hold or something, OK?
I think if I do this, I might go into that.

Tom Keel 20:44
And you got a question.

Gianni Vesuviano 21:00
OK, so I think finally we got one organiser in each of the rooms, and also you, Tom, in the main room.

Tom Keel 21:07
Fantastic, cool.
Yeah, like I said, there's a there's a bit of text at the at the top of the notebook, so I expect that maybe it'd be a bit quiet for the 1st 10 minutes.

Hopefully.
People will interrupt a little bit.
And then Gianni, obviously I can't see the rooms either. Would it be okay to like push me into them one by one? Is that going to be awkward to do?

Gianni Vesuviano 22:07
You're an organizer, so you'd have to move yourself between them one by one.

Tom Keel 22:11
Ohh, I see.
Maybe I should...
I mean, they might in those laptops.

Gianni Vesuviano 22:19
I can see all the rooms chats and so far nobody said anything except Kit who said please say if I can help.

Tom Keel 22:30
Okay.
Oh, actually, I see these room chats.
Okay, Gianni, so if you stay in here, I'm gonna try to join room one, have a listen and then go through.

Gianni Vesuviano 22:51
Okay.

Tom Keel 31:51
All right, yeah.
If you don't mind.
Yeah, I mean, if you don't mind keeping it on your personal laptop, that's obviously completely fine. I mean, the main takeaway is that, well, yeah, you can run these locally. That would be the best case scenario to do anyway. And then we'll release sort of the answers later as well. But then you can always come back to, because there's aspects of this workshop which we probably won't finish today.

So, but um, yeah, what is it that very first PIP install has failed? Is that is that what you mean?

Gianni Vesuviano 32:55
Tom, when you left the main group to go into one of the breakout rooms, your screen stopped sharing.

Tom Keel 33:04
Yeah.

Gianni Vesuviano 33:05
Could you please put that shared screen up again?

Tom Keel 33:12
Not yet.
That won't be a problem. Sorry, I need to, I'm accidentally in the main room as well.
So Gianni said something in that.

Tom Keel 33:24
Gianni, can you?
And...
Hi, Gianni, are you still there?

Gianni Vesuviano 37:43
Yes.

Tom Keel 37:45
Sorry about that. I...
Yeah, I don't know what's happening. I think it's Teams Linux just breaking everything. Was I coming through like with the microphone on this chat as well?

Gianni Vesuviano 37:58
At certain points, yes.

Tom Keel 38:03

Okay, that might be a problem. Have you got admin to mute me or haven't you when I'm in?

Gianni Vesuviano 38:10
Yes, I'm able to mute you.

Tom Keel 38:12
Okay, I'll just, I'll put my hand up or something like that. If I'm, if I think I can, if I can hear you, yeah, I'm having a conversation with someone else. Because it's a bit distracting. And just mute me on this chat. Yeah.

Tom Keel 38:28
Sorry about that. That's just, that's Ubuntu, I think. Well, Teams.
Yes, just to reiterate, Moa'az, if we get to, if we do section one and section 2, don't worry if you finish all of that. But then I'll go through after we have a break at sort of half past two, well, so 40 past two, I'll go through for 20 minutes and go through the entirety of those two sections.
But I'm happy to answer questions on any of it now. That goes for everyone in the group, in the meeting right now.
Like, it'd be interesting to hear sort of how people used Python and Pandas before, or is this sort of the first time they were using it?

Tom Keel 40:19
Yeah, we've got 5 minutes left on the original timer. OK.
Oh, fantastic. Yeah, so that would be good for, hopefully today we can lean into some of the things that might be good to set up, like if you were going to turn this into like a paper or something, like how to set up like code in a way that can format your data and then run a model.
Also, I think tomorrow will be more interesting to you because we're doing more machine learning stuff as well.
There's no more questions.
Oh, yes, and Joshua, I see you've done RStudio and QGIS as part of this course. Yes, I actually was talking to someone in room one about how the TIFF file that I use for the orography data was sort of extracted using QGIS, because essentially
I think it's the Ordnance Survey potentially have topographical data and then that

needed to be put onto these regular grids. It's the same grids as the rainfall data. So I think, yeah, with all of these workshops, we're just going to use, yeah, just Python or just R.
But very much like for any sort of proper, yeah, modelling workflow, you would go between lots of them. Because different ones do different.
data formatting, like, so QGIS is, if you're working with any spatial data, yes, you can do a lot of stuff in Python with this GeoPanels library, but you're not going to be able to drag and drop things and you're not going to be able to interact with it in the same way as you would something like QGIS. Yeah.
It's good to hear the lesson experience or not.
Yeah, Joanna, I reckon we could bring people back. Or if there's like a 30-second countdown or something, I don't know.

Gianni Vesuviano 44:01
Yeah, I don't know how to do the countdown, so should we just wait 30 seconds and just close the rooms?

Tom Keel 44:06
Yeah, sure.
Yeah, yeah. I'm sure people will be in the conversation.

Tom Keel 44:15
And yeah, Joshua, it's completely fine to do. Yeah, whatever you're comfortable with, just do it on that thing. I mean, yeah, I'm stubborn about sort of what was going on about before for viewing NetCDF files. I was still using ncview, which is a command line tool, which is just horrible for doing stuff. It always happens.
Yeah, and Gianni, I don't know if you saw that one of the people had to leave, so...
Be one less person.

Gianni Vesuviano 44:51
Yeah.
Okay, it's 2:20, so I'm going to close the rooms.

Tom Keel 44:55
Yeah, fantastic. Cool.

Yelesen, Sebastian 45:10
I think that's not so much of A concern.

Kit Macleod 45:13
No, right.

Jean-Yves N'guessan 45:14
Ohh, time is over.

Kit Macleod 45:17
Thanks, Sebastian.

Yelesen, Sebastian 45:19
No problem.

Tom Keel 45:22
So, welcome back everyone. I think we all should be back in the main group now. Oh, I see Matt is still on the meeting.
It's just the next thing, I think.
Oh, he's back. You back, Matt?

Amulya Chevuturi 45:38
No, Matt is back. Matt is back.

Tom Keel 45:39
Okay, nice.
Yeah, my laptop almost set alight, I think, with the amount of jumping between rooms. That was quite a lot of processing today.

Matt Dalle Piagge 45:46
Mhm.

Tom Keel 45:53
So yeah, I just opened this up for, my timer's just gone, which is good. So we'll have

our break at half past, maybe we can finish a few minutes early. Could I just have a show of hands for anyone that's completed all of section 1 and section 2 in that time? I don't really know if the timing is good for that or not, but just raise your hand if you're there.
Did see Sebastian? Yeah, that's smart.
But yeah, that's completely fine, of course. So I will go through all of section one and section 2 in the 20 minutes we have after this 10 minute break at half past. Actually, yeah, let's just get.
Oh, so share my screen, just the schedules, um.
But yeah, was there any outstanding questions before we have that break in, yeah, as I said, 5 minutes or so? This is the...
schedule. So yeah, we've just in that final bit of finishing practical before the break.
Or Amulya, how did you... Oh, I see you're sneezing there. Sorry.

Amulya Chevuturi 47:09
Sorry.

Tom Keel 47:11
How did you get on? Because I saw that you were doing it.

Amulya Chevuturi 47:14
It was great. I think I was doing along with the participants, so I was working through it. And Victor kindly helped Adi and Martha with the query they had. So basically we were all sorted, well, other than me. But it went great.

Tom Keel 47:20
Okay.
Fantastic, yeah. Thank you, Victor, for doing that.

Víctor Malagón Santos 47:35
No problem.

Tom Keel 47:40
Yeah, so, um...
Yeah, so that there was a there was an optional task in there as well.

But.
Yeah.
If yeah, don't worry about.
Wait, where was the essential task? So there is, there's one thing in section 2.1 which is called an essential task. Oh yeah.

Amulya Chevuturi 48:03
Tom.
Sorry, could you make it bigger? It's 60% currently. I can't. They're very tiny.

Tom Keel 48:12
Yes.
Is that okay?

Amulya Chevuturi 48:18
Yes, thank you.

Tom Keel 48:20
I completely forgot I was sharing my screen actually, but yeah. Yeah, so yeah, so this is the only one that is sort of essential because we use this, these two variables later on. I will go through and run this, but this is just when we come back from that small break. This is the one that you really need to
have these question marks filled in properly. Otherwise, something will break later. That's also to all the code conveners. Yeah, so just come back to this task if you get an error with, yeah.
On.
The.
But yeah, I'm still playing this.
Yeah, I'll just come back and do this.
I see that, uh, there's Augustine still in on hold, don't know that's still in a group or something.

Gianni Vesuviano 49:25
Nobody is in the breakout rooms. They're all closed. So the on-hold will be unrelated to that.

Tom Keel 49:27
Oh, yes, this.

Tom Keel 49:30
Sure, yeah.

Amulya Chevuturi 49:31
There's a lag with the rooms, just ignore it.

Tom Keel 49:34
Yeah.
Well, I guess we could just, I might just say we go on break now then, and then I'll come back and run through it. Come back at 22, so I'll leave it on this on this slide so we can see.
Um, we'll come back at twenty-to.

Amulya Chevuturi 50:01
Martha.

Tom Keel 50:03
Oh, is there a question?

Amulya Chevuturi 50:03
Martha has some questions.

Martha McKee 50:08
No, sorry, I keep pressing it by accident.

Tom Keel 50:10
Oh, sure, yeah.

Martha McKee 50:11
Don’t worry.

Amulya Chevuturi 50:13

So we are coming back at 2:40.

Tom Keel 50:17

Yeah, 2:40. Please only spend at most another 5 minutes. Please step away from the computer for at least 10 minutes before you come back. That's your task. All right. See you then.

Seemed like a very long time, then.

Ohh.

Matt Dalle Piagge 1:03:47

It was a proper break.

Tom Keel 1:03:49

Yeah.

Okay, well, we'll get started again.

Just assume that people are pretty much back or they're just about to be back.

Amulya Chevuturi 1:04:10

One more minute.

Tom Keel 1:04:11

Ohh.

Ohh yeah, got one more.

Tom Keel 1:04:45

Thanks.

Okay, so we've got another 50-minute block of work, 20 minutes where I'll just go through everything we should have done up until this point and then we'll start section 3. So as

I put on that slide, but we are doing the very first bit, which is loading in rainfall data and then formatting that data before we then what we're about to do after this, which is then start to compare rainfall data from different sources.

So I hope you got a chance to sort of read through a bit of the introduction,

although I did sort of, in that intro outline, went over some of this.
Yes, so these are the different data sources we're loading in. So we're loading in that CEH-GEAR data set, so gridded estimates of areal rainfall, which is produced at the UKCEH. This is from JASMIN Object Store, which we were talking to you before in Workshop 3.
We're also getting rain gauge data from that same, the object store, the clouds, cloud-based data. We're going to be getting some sub-catchment boundaries. Again, Matt went over this in his practical. So these are just the outlines of the Severn. Specifically, we're getting a few of the sub-catchments of the Severn.
that are up in Wales in the upper parts of the Severn. We're also getting water course data, so essentially just the water, the river, and the tributaries of the Severn region from OS open rivers. So this is all like publicly available British, UK data, basically.
We're then using a similar bit of code to get that, well, in this case, the GEAR daily, so that's daily CEH gridded data, and they're in 100 kilometre chunks. So if we have a look again, we'll see, oh, this is in.
It's not in spaghetti format or lasagna format, but that's something else.

Matt Dalle Piagge 1:07:17
What are we gonna call that then?

Tom Keel 1:07:17
I don’t know.

Amulya Chevuturi 1:07:19
The whole log.

Matt Dalle Piagge 1:07:21
What?

Tom Keel 1:07:21
Okay.

Amulya Chevuturi 1:07:22
Wasn't this the ravioli?

Matt Dalle Piagge 1:07:24
Oh yeah, that was it. Yeah, yeah. Ignore us, guys. It's fine.

Tom Keel 1:07:26
Okay.

Amulya Chevuturi 1:07:29
Sorry.

Tom Keel 1:07:31
ravioli format chunking. But this is daily data. But again, do not load this all in. This is 310 gigabytes of data and every individual chunk is 27 megabytes by itself. Oh, damn, this is 50 seconds. Sorry, we'll have to wait off to talk for a bit while that loads. But essentially,
We're going to then plot out the rainfall data from that. So again, as we saw in workshop three, we'll just have a look at one day. This is that very first step that I said, like have a look at your data before you start going further and further. What does it look like? What sort of things might you be able to do with it? And just simply having a look at data gives you an idea about
where you might go with the research. And importantly, for stuff like if we were when we were plotting, when Matt was plotting like the catchment data, there is also gaps in it that you can see when you plot outspoken data.
So, we load the same.
Then we will be getting the hourly gauge data, again from the object store that we have as part of the flood and droughts research infrastructure project at the UKCEH, and we'll get some metadata for that as well.
So, we work that in.
There you go, brilliant. It took one minute, not 50 seconds. So yeah, so this is just rainfall for that one day in the 13th of February. We then look at the data that we get from our read CSV. We're reading this file, which is stored in this S3 bucket called rain gauge. It's quite so importantly, we're using PL.
instead of Pandas to read. So something that, well, we were actually talking about at lunch is the fact that Pandas doesn't isn't explicit about what data type different columns are in. It will often try to work it out, whereas Polars is very strict about

Okay, so in this case, we have the CSV, the ID column is an integer, which can have 64 places of numbers in it, basically. And date time is in nanoseconds and precipitation and rainfall basically is a floating point number with a maximum of 64 decimals as well.
So yeah, so this is hard type basically.
things. And then similarly for our metadata, we have an ID column, we have, in this case, ignore the source ID, we'll be using the IDs rather than the source ID. This is the ID as it's called on their own database. So because we're looking at a region of Wales, it'll be Natural Resources Wales, and this is their ID that they use for this.
This first one, the Strongford W WKS. A location, so we use the OSGB.
location, which is in metres for the UK, and an elevation that is up. So we can have a look at a subset in the data. So we'll just select one of these gauges.
So we selected ID of 90537. So what the data looks like, and it runs from 1961 to 1989, so it's not actually a gauge which is still in use. So the first task was very simply, can we subset the precipitation data to the station ID 904?
So you just simply copy and paste.
that code in because we're asking the code to filter our data to the ID 90492.
Fantastic. We've got precipitation data for that station now. So this station ran from 1973 to 1988. Again, it's no longer being used.
Uhh...
Have a look, we can sort it to double check that.
The date time doesn't go past.
Yeah, so it is, so it's 1973 to 1988.
Then there was an optional task. So we'll just use the same methods again. So we want to filter our metadata and we want to have a look at the column which stores the name authored. So I'll just scroll up to have a look at what the metadata looks like. So it looks like the names of the stations are stored in this column name. We will put the word.
Name in there, and we will ask for the exact string.
Okay, fantastic. So we see that this is ID 42931, so we can filter that. In fact, we could even get fancier and just plug this directly in.
And say that we want this value and we call this Wolford ID.
So now we've got a variable that has
The number that we need, we can use once again, so rather the pandas, we would do this with lock. We're doing this with the filter methods, which is the main thing I'm

going to be teaching you for.
Running.
Toiler's codes say this is that fantastic. So this is just an example of how the filter methods works more than anything else. We're then going to go and get some catchment data. So this is again from the NRFA, the catchment boundaries and the ordnance survey is going to be where we're going to get that.
watercourse data from.
You can then plot this.
So, we're reading in these files using GeoPandas rather than from the object store. These have an exact location, so we can go to the folder and see that when we've done this, wget is essentially going to the internet and downloading data. All of this catchment data is stored here. So, these are the names of the different catchments. So, Abermule, Bewdley, Dolwen. I don't know why there's extra ones.
And then we can then plot this data using just essentially loads of these geopandas dot plot methods, basically. So plot the shapefile for Abermule, which I believe is this big outline of shapefile, then view of lead, then dull and then.
And then we also want to plot, again, we've got red this rivers line string in as a geopan is going to plot that river through it. So that's just a way of representing, here's our study area. And if you were to pick up a different region of the UK to look at, you could say, okay, what are my sub-catchments? What state files do I need? what room loads do I need? And there we go, you could repeat this for that. And what do the outline borders on this plot represent? If anyone wants to just quickly chime in with an answer.

Yelesen, Sebastian 1:14:33
They represent catchment boundaries.

Tom Keel 1:14:36
Fantastic. Yeah, that's exactly it. So I appreciate that we looked at this in the last one as well. So just to remind, that's what that plot is. This is our study area. So when we use this catchment boundary later, we're going to say, can we subset rainfall just to this catchment boundary?
In each, in each case, can you subset it to each of those boundaries?
We're then going to load in some data using Xarray and Raster IO. So Rio Xarray is basically a Xarray third-party library, which is built on Xarray, but for reading in things

like TIFF files.
So if anyone's got any experience working with spatial data formats, TIFF is another one. It's basically just spatial data. So we want this basically because, as I mentioned at the start of that opening thing, the region is quite topographically diverse. So there may be some research questions that can go down saying, well, does it rain more in areas where
The grid cells are at higher altitudes, for example, or are we less certain about those grid cells?
Okay, brilliant. So that's data loaded in. We've had a look at it. We know what we're dealing with. We've got four or five different sources of data.
Yes, we're going to format data. So this doesn't matter whether we're doing environmental data, financial data, blah, blah, blah, blah, blah, blah. It is really important that we do cleaning and formatting of our data. There's always outliers, there's always things that we need to use our own domain expertise to subset. And I think I'll do a very simple example of
Let's remove stations from the rain gauge data that are particularly gappy or that don't extend very long periods of time.
will just take data from 1990 onwards. So remove any gauges that start, have a date time before 1990.
This is essentially asking for filter the rain gauge data where the date time is more than 1990.
Let's have a look at one of those data to see what it looks like. So, you may not know if you're not familiar with working with things like matplotlib, but essentially when you plot a time series and it has gaps in it, it will automatically interpolate to the next point. So, these lines don't actually represent data, they're just an interpolation.
But if you imagine they weren't there, there's huge periods of time where there's no rainfall coming out of this rain gauge. So we probably just don't want to put this into our final estimates. This is something we don't want to put in our data at all. We can do this by looking at completeness. This is quite a normal measure to remove things like, well, given how long the period of rainfall was for a given gauge, how many days did actually have rainfall in them? This just basically gives you each gauge, so you're grouping by the gauge ID, start and end dates, and then tell me how long The.
the number of rows are. So this gives you, okay, so for this gauge, there's 1990 to 1998 and there's 2517 days. So for eight years, is that enough days? We don't know.

So we'll look here and we'll look for a, we calculate an expected days basically. So This is just doing a date time comparison. So this is a little bit more involved, but essentially because we have a date time object for end date and a date time object for start date, it is going to calculate a time delta between this date and this date. And it's very nicely got a dot total days. So we can just use that as our number of expected days.
So look at this again, we'll see, right, so for this very first one, we have 2,517 expected, but there should have been 3283. So that means how many gaps there are. We can then divide one by the other to get a quick percentage. So again, for this first one, it's about 76%.
Complete, so there's there's quite a few gaps in that, and that one probably will remove.
We can look to filter our data. Oh, in this case, we're looking for data that's particularly gappy.
Is that do I need to go back? Amulya is that? I've not seen the chat.

Amulya Chevuturi 1:19:10
I have posted it.

Tom Keel 1:19:12
Oh, fantastic. Cool.
Yep, and then we can also plot out, this is another way of getting a view of our data. So for each of these individual rain gauges, at a given point in the Severn catchment, how many of them are like, what's the distribution of completeness? So it looks like, I mean, I arbitrarily pick 90% in the next essential task for us to do that. I think we can essentially throw out all of these because we don't have a good enough rainfall record for them.
But you could easily just draw that out 70%, 80%, it doesn't matter. Okay, so this is the part that we need to all do and then I'll let you get performance to the independent task. So we have been asked to filter out the data so there's that word against the filter with less than 90% completeness. And so essentially we're doing this because we want our data
to be only these quote unquote reliable gauges. So what we can do is we can set an ID called reliable gauge IDs. We say we want to filter. We want to use the percentage completeness column, which we've created up there. So just to show you what data

we're looking at. So this is our data. We've got ID start and observe expected percentage completeness.
The last few steps have gone through how to produce those, and we want to filter based on this percentage of completeness, so we want to ask for 90%, so 90, and we're going to essentially look at that as well. So, just have a look. So, this gives us a list of all the gauges that are above 90%.
Fantastic.
Brilliant, so that's the very first step. So I think there was only two things to change. It was to add percentage completeness to here and 90 to here.
And then secondly, we want to filter only those IDs as two steps to 90% completeness. So again, we use our filter. This time we want to filter our rain gauge data. So remember, that's the data that has literally just ID, precipitation, date, time, 3 columns. We want to say the ID has to be in this list of gauges that we said were reliable.
So.
Yes.
And then we'll have a look.
See if there's any data in there whatsoever. OK, fantastic. So we have some data in there. Just have a quick comparison to what it looked like before. So we've got 348 rows, 3 columns before we had.
381,000, so we've removed that 40,000.
I realise we're only on 2.2, so go through these a little bit quicker now, we get 3 minutes left. And we then want to subset the data. So we've loaded in that CEH-GEAR data, and as we saw earlier on, we have got data for the entirety of the UK. We just want to subset that so we take this small region in here in the middle of Wales in the Cambrian Mountains we just have Severn.
Where are the two parts?
It's 2.2, yeah, should take 45 seconds.
So, we're essentially doing something with Xarray where we're saying we want to use this Easting range, which we set right at the top, and this Northing range, and we want all the data from 1990 onwards.
And this stuff is essentially just to speed up the processing. So we have a variable in there called minimum distance that calculates the minimum distance from a grid cell to its nearest gauge. Don't need that for this analysis. So before we load in this data, which is on the cloud, we'll just remove that and speed the data loading up.

And then we looked at sort of seconds.
Oh, we're at 3, yes, just one time.
Loads.
Yeah, so we load this data in. This is an Xarray data set.
and we'll plot again those shapefile boundaries using these GeoPandas using that plot function. And we're also, because this data is in that tabular format, we'll draw a scatter plot. So what this will do is give us all of our gauges and all of the, sorry, all of the subcatchments and all the gauges within those subcatchments.
Yes.
We've already seen this figure, 'cause I used it in the watch overview.
But yeah, we're just about to send you off for another half an hour block to do independent work, starting from three. Was there any questions before we do that?
It's not.
Gianni, how ready are you to send people into the breakout rooms again?

Gianni Vesuviano 1:24:33
If they're going into the same rooms, then I can do it right now.

Tom Keel 1:24:39
I don't see why not. I mean, is it awkward to get it randomised or is it just easier?

Gianni Vesuviano 1:24:45
Uh, let's have a look.
Yeah, I can. I've reshuffled everybody.
I just need to take everyone out of room 4.

Amulya Chevuturi 1:24:58
When do we come back, Tom?

Tom Keel 1:25:01
Yes, so that's taking, it's been 2 1/2 minutes now. I hope this isn't a problem for everyone. If it is, I'll put a quick fix in the chat. But when I run this before, I needed 45 seconds. But I'll get the schedule back up. So we're going to continue now for the next half an hour and then we'll have a break and then we'll go over

Tom Keel 1:25:20

in a similar fashion if it's useful, section 3 before we get onto that independent task 4. So you have.

Yeah, 25 minutes in the breakout rooms and then we'll bring you back for any burning questions, but other than that.

Yeah, I mean, is it actually a very quick show of hands? Has anyone got to this point and it hasn't loaded because it took 45 seconds when I tried it before?

Tom Keel 1:25:49

Okay, you have, okay, nice. It might be loads of people trying to access the same resource. Matt, do you think that's likely?

Amulya Chevuturi 1:25:56

I think your Google might have given up.

Tom Keel 1:26:00

Yeah.

Matt Dalle Piagge 1:26:00

It's possible. Who knows, to be honest.

Tom Keel 1:26:02

It took a while, okay.

Amulya Chevuturi 1:26:02

Yeah.

Matt Dalle Piagge 1:26:04

But yeah, it only took 45 seconds or so for me.

Amulya Chevuturi 1:26:08

Yeah.

Tom Keel 1:26:09
Yeah, what worst case scenario?

Amulya Chevuturi 1:26:10
You might want to run a refresh. Yeah.

Tom Keel 1:26:15
Ohh, it ran for you as well Gianni or is that something else?

Gianni Vesuviano 1:26:18
No, I got another.

Yelesen, Sebastian 1:26:18
Yeah, mine.

Tom Keel 1:26:21
Open.

Amulya Chevuturi 1:26:21
Gianni has a comment.

Tom Keel 1:26:26
Uh...

Tom Keel 1:26:27
Yeah.
Oh, Sebastian's took 33 seconds. Robyn, is that a question or just hands up from this one? Oh.

Jones, Robyn (MECP) 1:26:42
Question, is there a way to like, because I'm on like 3 minutes as well for it to load, is there a way to like stop it and restart it or does it just...

Tom Keel 1:26:44

Yes.
Yeah, so I think it is a internet thing because everyone's trying to access the same resources. So it's not necessarily frozen. It's more likely it's just waiting for your turn in the line to access the resources. I mean, mine's just run for 4 minutes, 10 seconds. The actual.

Jones, Robyn (MECP) 1:26:54
Keep going.

Tom Keel 1:27:14
like time it took to run that code wasn't very long at all. So it's just an internet problem I think. Leave it, if it gets too closer to like 10 minutes, maybe you can come in and change how you're subsetting it. So here we're asking in this line here, we're asking for 1990 onwards. You could
just take a really small subset, go like 2005 or something like that, that will massively reduce the amount of resources you're accessing. But hopefully just wait a few more minutes and then it will work.

Jones, Robyn (MECP) 1:27:38
One second.
OK.
Do we need that to do part 3?
Listing.

Tom Keel 1:27:51
Oh no, yes, you do. Yes, so we will be using this again. So yeah, let it run, ideally. And like I said, just change this line if it's still a problem. Change it to 2005.

Jones, Robyn (MECP) 1:27:53
Yeah, OK.
Okay, thanks, Tom.

Gianni Vesuviano 1:28:07
Right, I do actually have a question.

Tom Keel 1:28:07
Okay.
Yes, please, yeah.

Gianni Vesuviano 1:28:10
All right, because Kit has now left, we've gone from four workshop leads to three, so should I reduce the number of rooms to correspond with that?

Tom Keel 1:28:21
Go on then, yeah, why not? Because I will go around each of the rooms again, but my teams might crash, so I might be a bit slow doing that.

Gianni Vesuviano 1:28:32
So there'll be 3 rooms then for the three workshop leads.

Tom Keel 1:28:37
Fantastic, yeah.

Amulya Chevuturi 1:28:37
Who are the three workshop leads?

Tom Keel 1:28:40
Ohh.

Gianni Vesuviano 1:28:40
Well, you, Tom, you and Matt.

Amulya Chevuturi 1:28:42
No, to like.
But no, okay, yeah, whatever. But then Tom will leave the session there, his work.
Okay, fine, that's fine.

Gianni Vesuviano 1:28:49

Yep.
So, should that be two rooms then?

Amulya Chevuturi 1:28:55
Tom.

Tom Keel 1:28:55
Do 2 rooms, that's fine a bet, but make sure.

Gianni Vesuviano 1:28:58
Two rooms.

Tom Keel 1:29:00
Yeah, again, I want to see some interaction between everyone and yeah, cameras on.

Gianni Vesuviano 1:29:09
OK, so in that case, there'll be the main room and one breakout room.

Tom Keel 1:29:16
Ohh.

Gianni Vesuviano 1:29:16
And...

Tom Keel 1:29:18
Do, no, do 2 breakout rooms and then the people that can't join the breakout rooms stay here.

Gianni Vesuviano 1:29:25
Oh, okay.

Tom Keel 1:29:27
Sorry, Gianni.

Gianni Vesuviano 1:29:32

All right.
So you, there are 19 participants and six of them are not able to join a breakout room.

Tom Keel 1:29:44
Yeah.

Gianni Vesuviano 1:29:45
So.
You want all 13 of them?

Amulya Chevuturi 1:29:49
They've...
They will be with you in the main thing.
Gianni.

Gianni Vesuviano 1:29:56
I think I'll just, I'll just split it 10 and 9, 10 in a room, 9 in the main.

Amulya Chevuturi 1:29:59
Yeah.

Tom Keel 1:30:01
Sure, well, that's fine, yeah.

Gianni Vesuviano 1:30:01
And...
Let's go. I'm opening room one.

Tom Keel 1:30:13
Oh.
Turn it up.
Okay, cool. I'll set a 25-minute timer.

Amulya Chevuturi 1:31:21
Not sure why Mayuran is being congratulated, but you go there.

Tom Keel 1:31:26
Yeah.
It's helpful.

Amulya Chevuturi 1:31:31
Wait, I'm so confused. Will there be another breakout room or am I just staying in this room?

Tom Keel 1:31:36
I think you're just staying in this room, it sounds like. Yeah, so everyone that's in this room.

Amulya Chevuturi 1:31:38
Okay.

Gianni Vesuviano 1:31:40
You're in Breakout Room Zero.

Amulya Chevuturi 1:31:42
Right.

Tom Keel 1:31:42
Yeah.
Yeah, so everyone in this room, feel free to go cameras on and have a chat. We'll try to remain quiet-ish. Or if there's any questions, but yeah, just working on section 3 now. You can skip over if you've not finished section one or two. There's only that essential task that Mayuran has kindly shared.
That we need to have run to do section 3.

Yelesen, Sebastian 1:32:22
Hello?

Tom Keel 1:32:24

Hello, yeah.

YS **Yelesen, Sebastian** 1:32:25

Hey, for block 54.

Should the should the rain gauge plot?

So this is where you plot the nearest grid cell below the rain gauge ID, the Montford rain gauge.

Tom Keel 1:32:44

Ah, yeah, so where is this? So section 3.1 is that?

YS **Yelesen, Sebastian** 1:32:45

And.

As well.

It's Block 54, so you almost...

Tom Keel 1:32:56

Yeah.

YS **Yelesen, Sebastian** 1:32:56

It's why you plot the two plots. I haven't had any errors executing the blocks leading up to this, but the red, the nearest grid cell one has present has data, but the rain gauge does not. So it's not, well, it's not plotting it.

Tom Keel 1:33:04

Yeah.

Ohh.

Yeah, I mean, it's not going to be block 54 because this is just about the run order.

So.

YS **Yelesen, Sebastian** 1:33:21

Ohh, sorry, right.

Tom Keel 1:33:23

Is it, do you mean this one?

YS **Yelesen, Sebastian** 1:33:26

Yeah, that's it.

Tom Keel 1:33:27

Okay, fantastic. So you're saying you've got rain gauge, but no.

YS **Yelesen, Sebastian** 1:33:28

So I haven't had, yeah.

Tom Keel 1:33:32

Interesting.

YS **Yelesen, Sebastian** 1:33:32

I've got no rain gauge, but it's acquired the grid cell observations properly, and I haven't had any errors leading up to this point.

Tom Keel 1:33:44

Okay, right, yes. So if you were to plot, so this is the gauge that we're plotting. So essentially this variable here is storing the data. So if you were to plot that gauge, does that have any data in it? Because I think it may be related to your other point that you put in the chat.

YS **Yelesen, Sebastian** 1:33:50

I'll do what you've done here, where are you?
One gauge.
Right.
So.
Press here.
Code one_gauge.
That's odd.

Yelesen, Sebastian 1:34:38
So yeah, when I, when I, when I try and print, it's the shape of the one gauge data frame is 03.

Tom Keel 1:34:46
Zero 3, yes.

Yelesen, Sebastian 1:34:47
So, yeah.

Tom Keel 1:34:49
So that is a problem with this then. Is it... have you plotted the filter? Yeah, that that simply just goes back to I think.

Yelesen, Sebastian 1:34:53
Yeah.

Tom Keel 1:35:01
That was that essential task. Yeah, so can you just double check that these look similar to this, that you've got your reliable gauge and that is actually has some data in it.

Yelesen, Sebastian 1:35:09
Hey.
Sure.
Yep, step one.
pl.col percent completeness.
You won't believe this, but I got everything correct except the direction of the inequality is the other way around.

Tom Keel 1:35:42
Oh, right, yes, yeah, no worries.

Yelesen, Sebastian 1:35:46
Yeah.

Tom Keel 1:35:48
Yeah, that's annoying that error because it's...

Tom Keel 1:35:52
Yeah, it doesn't look like it's an error, but then there isn't any data in there with the subset.

Yelesen, Sebastian 1:35:58
Let's see if, uh, if that's...

Tom Keel 1:35:59
Yeah, you should be able to jump right back to this cell now.
Where you're setting one_gauge equals that filter.

Yelesen, Sebastian 1:36:09
What's.
Can I have a look at your reliable gauge ID block again, because it does look it does look consistent, but I might have missed.

Tom Keel 1:36:28
Yeah, sure.
There we go. Yeah, I'm gonna post it in the chat as well.

Yelesen, Sebastian 1:36:33
Yeah, yeah.
Yeah.
That's right.
Put completeness greater than 90.
Yeah.
dot filter Severn rain gauge data.
Completeness. Yeah, that looks fine.

Filter call ID is in reliable gauge.
Those that looks that looks correct.
So.
Why is networking then?

Tom Keel 1:37:27
Does this data print anything? Is that you're still getting 0 here?

YS

Yelesen, Sebastian 1:37:33
It's got Severn rain gauge data filtered.
How will I print the heads of that?
Yeah, it does. It does have observate. It's got, yeah.

Tom Keel 1:37:54
Instead of the heads, just get rid of the head arguments so you can see the shape of it. It's possible that you're not getting 348,000.

YS

Yelesen, Sebastian 1:37:54
Understood.
Yeah, yeah.
No, that's fine. I've removed the head part and the shape is now consistent with what you've got.

Tom Keel 1:38:09
Oh, brilliant.

YS

Yelesen, Sebastian 1:38:12
So I'll just try and run everything from then on.

Tom Keel 1:38:17
Yeah, yeah.

YS

Yelesen, Sebastian 1:38:18
And just see, see what changes.

Tom Keel 1:38:37
How is everyone else getting on with this? Is there any other questions you can use your mic or in chat?

Sam Tulloch (PGR) 1:38:43
90% of my problems were I kept spelling gauge wrong.
So.

Tom Keel 1:38:49
Okay, yeah, that happens, yeah.

Sam Tulloch (PGR) 1:38:51
Other than that, it's behaving itself.

Tom Keel 1:38:56
Is it the A and U the wrong way around or is it the?

Sam Tulloch (PGR) 1:38:59
Yeah, well, I've got quite bad dyslexia, so if it's not spelled exactly the way I like, I typed it out, like sometimes like I miss it or I'll type it wrong and I'll not notice. Usually when I code, I've got like, it's set up for the syntax goes every single colour when it's

Tom Keel 1:39:10
Yeah.

Sam Tulloch (PGR) 1:39:18
like, just so that it's easier to spot when something's not behaving itself because it's not gone the right colour and I can go, okay, there's a mistake there, but it's a lot harder in a notebook.

Tom Keel 1:39:26
Yeah, yeah, yeah.
Yeah, well, well, my advice would then be if you work with rain gauge data then just

call it station, yeah, because I do the exact same when I'm like typing this out, but always going.
It's annoying, wrong way around, but...
Yeah, the naming of things is...
Can be a real headache.
This is also quite a big block of cells to go through. So just to say to anyone, if you're at different stages, don't worry. Feel free to ask any questions about anything, even if it's really early on in section 3 or really late.
The.
This is a really tough optional task. I forgot about this one.
Hey, got still got about 10 minutes. I'm going to jump into the other breakout room.
So.
Please just, because this is the main meeting, please put any messages in the main group chat and I can always run back.
One second.

Amulya Chevuturi 1:48:03

I am here in case anyone needs any help. Tom says he'll be back soon.
I think the breakout rooms will be closing in 5 minutes. So if you start wrapping up, that will be great.
Tom has messaged and said, if you are still stuck on 3.1, please ignore the optional task, move on to 3.2, and just work through the cells that are non-optional. Thanks.
Two minutes remaining, guys.

Tom Keel 1:55:11

Okay, so hello everyone. I appreciate that that section is a lot more difficult and some of the tasks should be, should say optional tasks. But I will go through this after we have a short break. I think potentially we'll make this break slightly shorter.
So I'll talk for a few more minutes now and then we'll come back at 40. But essentially, we're going to be moving on to the
Shows me.
We're going to be moving on to the final task after this. So, but of course, we've not actually got, not many people haven't finished.
Section 3 yet, so that makes things a bit awkward. I mean, just a show of hands, has anyone finished the entirety of Section 3? Just run through it. Don't worry about the

optional task and task for that. Has anyone run through the entirety of that?
Nice, awesome Sam.
OK, brilliant. Yeah, there's a few people. That's good. Is anyone at is still hasn't haven't got past section 3.2? So there's either a stimulus section 3.1, 3.2, so if you put your hands up as well, so we know.
Well, essentially, I'll go through as quickly as I can those first few sections. And then when we come back, ideally around 4:00, we'll start that final task. And that's time for any final questions or anything about the code, or if you wanted to, you've got questions about, oh, I've got this data, how would I?

Tom Keel 1:57:07
create monthly mean differences from it, how would we create climatology, etc, etc. We can go through answers. But I guess, I'll tell you what, why don't we, I'll start going through the section and then we'll have a 5 minute break and we'll stop. So yeah, this section 3,
is we're very much just comparing this gauge, Preston Montford, which is in the centre of, well, the Montford catchment to the nearest grid cell for that from the CEH-GEAR grid data. This just happens to be the idea of it, and this is what the data looks like, so it runs from 1990 to 2024.
We earlier subset all of our range data, so it started in 1990. Here's the metadata for it.
We want to, the very first thing, is select the nearest grid cell based on the information we have from the gauges. So in this very first one, we are just getting data that is from the CEH-GEAR daily data set that is nearby.
So, that's a single grid cell of 1 kilometre squared, which is directly related to the nearest grid cell to this point. So, when we, but yeah, that's essentially that point, and then we're then comparing them side by side. The first thing we can obviously see is that the gauge network.
for this data is being updated a lot more frequently. I may have mentioned at the start, but yeah, like I think the data runs out in actually yeah, 2019 for the CEH-GEAR, so we can't actually use anything past 2020. But this is like the very basic comparison we could do between the two and you could probably spend hours looking at this.
It's not a very good way of
Looking at this sort of data, which is very fine. There's a new value for every day. We want something that's a bit more statistical.

So the very first thing we need to do is create a data set that has both our precipitation gauge data and our grid data. So this is essentially it. So for every day for this given gauge, we have a value for the gauge.
a value for the gauge and a value for the grids. So the grid is to more decimal places because it's been interpolated onto the grids. We can run a very simple correlation. How well do these two time series line up? They're very highly correlated, but as we can see, the actual values aren't exactly the same.
So they are varying together, but they're not. They're not perfect.
We also can just create a new column. So essentially what I've done here is I've said, we've got this data, one gauge joins. This is our join data.
We want to create a new column, so we can do that with columns.
We want to say...
We want a minus pl.col precipitation gauge.
and we want to take away the gridded. So this is just the difference. The word alias here is just used to, what should we call this? We'll just call this like, call this Tom's test now, just a group point. So this is how we can create a new column. You could imagine you could use this to, oh,
Ahh, this one.
There we go. So we got a new column. We could do anything with this. We could be like, OK, times that column by that column. We so needed to do this. This is how you combine columns. You'd never need to do that. That's just that one. So we do this, we create a difference between the two. We're going to drop values where they're Either the gauge is missing data or the grid is missing data, because we don't want to compare those two. This gives us the difference through time. When writing this practical, I thought that I was thinking with my climate science hat on, where I see a time series of differences, I want to see if there's some sort of
relationship through time. Oh, is the difference between gauges and grids because of some climatological thing going on? Probably not, but that's what I think. But basically, we can see that the difference between the grid cells has actually some weird sort of pattern. It does seem like there's more variance.
we can sort of analyse that. So the mean difference is pretty much close to 0. But it looks like if we look at just a random subset of everything before 2009 and after, that the variation of the difference is slightly higher, implying that there is some temporal component.
not really essential, but yeah. There was this optional task which I'll really quickly go

through. This was a particularly mean task to start off with. This is basically asking us, can we get the number of days where the rainfall difference, i.e. this value, has a magnitude of at least one? So the data looks like this, and this is what the data talking to the rainfall difference. The very first thing we can do is we can apply a filter and we can ask for the absolute, so that is remove the sign, of the rainfall difference. If that is more than one, that's what we want to filter out, basically.
Absolutely, those brackets in the wrong place.
So this gets us 67 days where the, for this given gauge, the difference between the grid and the gauge was particularly high. So in this case, yeah, this case, the gridded estimate is much lower, also much higher than the precipitation estimate.
And in this case, it's the other way around, basically. We can then apply a group by, that's essentially going to do a count through time. So we say group by of time, and in this case, it's asked for number of days each year. So we say group by time, so group these 67
rows by time, group by every year, and we want to give us the sum of the range of difference.
So, this is only for the people that did this, so very quickly worked out.
How many days?
group_by_dynamic
There we go. So this just simply gives us a count through time of each year, how many.
Well, it's not showing, that's fine.
How many days were above? Okay, I will share this code at the end as well. And we're also looking at things where we're doing the running total. So in this case, we've got the monthly sums. So this is saying that for 1990 January, the gauge total was 102.8. And the grid total was 102.99, so it's pretty close.
I ask what is the monthly mean difference. So here I just want you to go doc mean to have a look at the mean difference. So it is.
Yeah.
Quite close to zero.
We can block that through time. So there was a particular month in the 90s where there was a huge amount of difference, maybe something happened, there was a change over a gauge or something, maybe that's something that needs to be quality controlled out. And equally towards the end of the...
Thing is these similar errors between the nearest gauge and the grids.

This is another nasty task which I should have set as optional, but essentially what I wanted you to do here is to group, get used to using this group by. So you're grouping the sum of the gauge in the grids, say every six months, and you're calculating a difference from that.
So we want to go call this.
Six months.
m6_rainfall_diff
Learn to plot out afterwards.
And we call this one_gauge_joined_6m
But.
So all that is doing is in a roundabout way, instead of just taking the monthly sum, it's just doing basically what we did for monthly and applying a different temporal resolution. So if we now wanted to do, I don't know, something random like 8 months, that works. If we wanted to do 8 years,
That works, just gives us a different way of looking at the temporal difference of a data set every time. Let's put this back to six months for now.
Again, sorry about that one. That should be an optional task. Also, this optional task is just basically to rerun all your code again. Because I think the thing I'm trying to hammer home here is the idea that if you write your code in a certain way where you change one line to change the resolution and you change one line to change the ID you're looking at, when it's written once, you can just have
other bits of code talk to it or just you just change one line and it changes the whole analysis. So that's just if everything's running quickly, basically it just says go up here, change to a different ID.
No.
And just rerun that, basically.
But I can't off the top of my head. Cool. So that is just some things we can do with rainfall difference. We're then going to do this. So I really want to spend slightly more time on this. So we'll have a break for 10 minutes. Come back.
Uh.
We'll come back at 50 past and then I'll talk for 10 minutes more just to really hammer home what we're trying to do with this final task. Yeah, sorry about sorry for rambling. I realise because it's overrunning. I'm getting a bit overwhelmed, but yes, OK, so please come back at 50 past.

Hey, hello everyone.
Final bit now, and then we can...
Go relax. So I am going to go through section 3.2 and then you've got another half an hour to sort of finish up or ask us any questions. We're not quite going to get to the final section of this workshop. What we'll say is that
There'll be, you can have it as sort of like homework and we'll be available throughout this week or even after that if you want to contact us. If you've got any questions about like those specific tasks, they get into a bit more detail about like, oh, how we can actually use the data we've loaded in here to answer some research questions.
Well, that being research questions about the Severn. So I'll continue and I'll just assume everyone is back. So we'll do 10, I'll do 10 more minutes of going through and then we'll go jump into our different breakout rooms again.
Or about just the same breakout rooms? Yeah, pretty easy.
So we are going to do 3.2.
Essentially, this starts with a massive load of functions or big wall of codes. And all these are doing are things that we've currently been doing in 3.1. So whereas in 3.1, we had loads of different blocks of code that were sort of doing this, oh, how do we merge cells together? How do we calculate the mean differences through?
of them. And then, yeah, these all these methods, well, basically are repeating code done before, but now put into functions. Now, if we want to calculate the mean difference between a nearest rain gauge, we can just simply say, well, which gauge do we want to look at? And there we go. We then have
this one big function which we're defining here, giving it which gauge we want to look at, and it gives us for that gauge, here is the nearest grid cell to the gauge, and here is the difference. So we just have a quick look to check if this is working something else. So right at the beginning, we called our metadata,
rain_gauge_metadata
Let's have a look at.
This is very nice. Let's go Whitchurch here, so we pick this ID, we'll throw this into this into the function set.
Ohh, that's one that doesn't exist within the time series is just about this one.
Neither does that one. Let's have a look.
Things with us.

The reason these might not be in the data anymore is because we've subset data that's patchy and data from
1991. Has anyone got another one?
Come on.
Great, okay, we'll just keep as it was then.
It should work for other gauges if that gauge data is in our data.
Yeah, so we've got this function that given an ID, it will calculate range of difference. We can then get all of the, oh, here we go, we should use this. We can get all of the IDs that have some data. And we can say each of these gauges, we want you to run that function and calculate
the rainfall difference. So it's basically just going to create however many IDs this is. 50 different unique IDs in this region. We're going to say, give us the mean difference, the max difference, the min difference. Get all that data.
It's quite quick to do. We can then create a data frame out of that. And now we have a data frame that says, for these 50 gauges which have data, give us the standard deviation of the difference between the gauge and the grid, the mean difference and max difference. And this is a useful thing for sort of saying, oh, is there any gauges within this?
small catchment that have particular problems. So we can also view that using a heat map.
This shows us those same statistics, the minimum difference between the gauge, which is essentially a bucket, and the grid, which is essentially an interpolation of the rain gauge network, the mean, the standard deviation, and the max. So we can see that there's a gauge here that has quite a high max and mean difference. So some of these gauges we potentially removed. This is a summary of all of the gauges, all 50 gauges.
We can also look at the correlation, so I won't go through this now, but release the code at the end, but go look at how we calculated correlation before.
We can join these statistics to the metadata. So then we can plot the data so we can have a look. Is there any sort of statistical relationship between where we're getting a large difference between the nearest gauge and the grid? It looks like, yeah, there is. There's something going on.
Up here, we got a slightly different difference. Let's do that max difference statistic. We can have a look and see that. OK, so it's pretty normal.
that column is not useful. But yeah, so it's pretty random. It's that there isn't any

geographical relationship between whether a gauge is closely aligned to a grid cell on that. Seems to be quite random. This is the point here where the difference is very high, but then right next to it, one that's not as high. So
There isn't not there yet, there's no spatial pattern between this, so maybe we can throw this all into the model and feel fine about that. So, section 3.3 will go over how we can compute catchment mean rainfall. So, in workshop 3, Matt went over how to do this with some of his code, but that's just another method that we...
we put in here. But essentially what we'll be doing is we'll be saying, given this field of rainfall, give us just one of the catchments. And then what we can do with that small subset is we can say, give us a catchment mean, so dot mean of that, of that. Region.
So, I will let you get on. So, if you haven't already, just jump to 3.3, run this cell, and then we'll come around the breakout rooms again. So, Gianni, if that's okay to put everyone back into the breakout rooms.

Gianni Vesuviano 2:22:10
Same rooms as last time.

Tom Keel 2:22:12
That's absolutely fine, yeah.

Gianni Vesuviano 2:22:14
Okay, opening the room one now.

Tom Keel 2:22:54
Okay, sorry.
So, I see, yeah, we've got Sam and Sebastian in this group again. I see that you had finished all of section 3, is that?
Fantastic, brilliant, good job.

Yelesen, Sebastian 2:23:21
Yeah.

Tom Keel 2:23:23
Nice one. I mean, just feel free to go on to that independent task and we're here for

the next half an hour to sort of help you like how you approach those things because they're very open-ended. They're very much like, how would you answer this research question? And we haven't given too much guidance. But the answers are basically, within the code that we've already been running through, you can copy and paste a lot of that.

Amulya Chevuturi 2:23:49
Tom, would it be good, prudent to get people back 10 minutes before 4:30 so that we can just...

Tom Keel 2:23:56
Wrap up.

Amulya Chevuturi 2:23:56
Do a final wrap up, yep.

Tom Keel 2:23:58
Yeah, no worries.

Amulya Chevuturi 2:23:59
Cool.

Tom Keel 2:24:05
Well, I guess my first question would be if you wanted to do the task now, which ones are you each choosing, Sam and then Sebastian?

Yelesen, Sebastian 2:24:16
I'm having a little bit, a little read of the tasks now.

Tom Keel 2:24:19
Mm.

Amulya Chevuturi 2:24:37
In the meantime, if anyone else has any questions, this would be a good time to ask.

Sam Tulloch (PGR) 2:24:45
I'll probably go for #3.


Tom Keel 2:24:48
Oh, nice. Yeah. That'd be good.
So you got any, have you had, have you got any ideas, Sam, like what you might do to approach this or what are the things that are coming to your mind for?

Sam Tulloch (PGR) 2:25:07
So, I'm an R Studio girly, so I would have, and I'm trying to work out how I would rephrate, like, redo it in this, but I would probably go for...

Sam Tulloch (PGR) 2:25:21
Like...
I would want to organise it based on what counts as a flood event first.


Tom Keel 2:25:29
Mm.

Sam Tulloch (PGR) 2:25:30
Um...
And then...
identify what kind of threshold or what kind of events would specifically be for that.
And then I would personally
Probably.
I don't know, I'd probably Mann Whitney it because I just have a tendency to man Whitney test anything.
I like that.
But I will.


Tom Keel 2:25:58
That's a different statistic, is it? You compare? Yeah, OK, nice.

Sam Tulloch (PGR) 2:26:01

Yeah.
to try and see if two things can be explained as being from the same population or not.

Tom Keel 2:26:10
Right, yes, that's it, yeah.

Sam Tulloch (PGR) 2:26:11
Because it's the one they use for drug trials where you've got placebo in one group and then whatever the drug is in the other group. And it's like, okay, has the placebo, do we have a difference in like the mean and the difference between points?

Tom Keel 2:26:23
Yeah.

Sam Tulloch (PGR) 2:26:25
And if you've got a binary of is it flooding or is it not flooding, it does work.

Tom Keel 2:26:30
Yeah, yeah, of course.

Sam Tulloch (PGR) 2:26:32
Yeah, I'll have a wee think and see if there's anything else I could do with it from the stuff we talked about today.

Tom Keel 2:26:39
Yeah, yeah. I mean, you definitely, that's the correct way to approach it. And that, how do we define flood events is potentially the whole, the whole thing rests on that because there's a whole, that's a whole field, like loads of people research that because there's no clear definition. And people that do study this stuff will be a bit upset that I've just picked out.

Tom Keel 2:27:00
Random days, because that's not really how floods floods work, so yeah.

Sam Tulloch (PGR) 2:27:06
Yeah.

Tom Keel 2:27:09
Sebastian, you can do task two. Fantastic, yeah, yeah. So this is a really interesting one as well, because we basically have, we built this gridded data set from the gauges that you see, basically, the gauges that we've been access to. But then there's all this historical data set from Carreg Wen, which is at the very top of, well, where the source of the Severn is basically. And this introduced a completely new time series. So the rain, I mean, I've looked at this data already and the rainfall is pretty much wrong for that grid cell. So it's like, it basically means if you were to use the CEH-GEAR, that gridded product, or the very, very upper reaches of the Severn, Your estimate would be way off, so your difference would just be really problematic.

Yelesen, Sebastian 2:27:57
Yeah, it's actually something like this I've done before because like in my industry, we're worried about natural hazards impacting like a very local site, like in it like a very local site that's only one or two square kilometres. But

Yelesen, Sebastian 2:28:16
Often, you know, because the nearest observation series that's long enough is quite far away. So you often either have to resort to use, you know, some form of gridded obs or maybe reanalysis or something like that.

Tom Keel 2:28:24
Yeah.

Yelesen, Sebastian 2:28:35
you still have to demonstrate how close, like, whether or not the, you know, those those might be representative of this very specific site is supposed to be studying, so.

Tom Keel 2:28:51
Yeah, yeah, exactly.

Yeah, and the further you go back in time with all this stuff as well, the more uncertain it gets, I guess, because the further away the gauges are, so...

Yelesen, Sebastian 2:29:07
Do you guys do something, what's it called, kriging? You know, and you have geo, yeah, like geospatial, like trying to work out how the observations might change as you go further away from an available source and these sort of things.

Tom Keel 2:29:12
Kriging, yeah.
So we don't with our products, but we'd like to introduce that. Essentially, the CEH-GEAR was developed with nearest neighbour interpolation. So it's a type of like, yeah, okay. Yeah, so that was basically chosen because that was something that ran quickly. Because as you can imagine, the amount of data

Yelesen, Sebastian 2:29:31
Yeah.
Yeah.
Yeah, yeah.

Tom Keel 2:29:46
And also when it was written in original implementation was in 2016, I think. But essentially what we will be doing as we move to higher resolution data would be using kriging with external drift, KED. So we just have a better estimate of the rainfall with an uncertainty term as well.

Yelesen, Sebastian 2:29:46
Yeah.
Yeah.
Mm.

Tom Keel 2:30:07
But as you can imagine, if you used it in your workplace, the amount of compute time absolutely skyrockets when you do that. It's not something you can run in a day

anymore. It's something that's over multiple days to produce the entire data set. Because it's just very, very computationally heavy thing to do.

Yelesen, Sebastian 2:30:28
Yeah.
It's a shame all these AI companies are wasting all this lovely compute on generating fake videos instead of using it on an actual good applications like those.

Tom Keel 2:30:36
Yeah.
Yeah, I know.

Amulya Chevuturi 2:30:47
Just to add to what Tom mentioned, GEAR, yes. GEAR also has like a secondary variable available, which does tell you like how far the nearest, like, you know, the data points are and those kind of things. I forget what the variable exactly was.

Amulya Chevuturi 2:31:24
But I do agree with you Sebastian that a lot of the compute power is being wasted. On useless items.

Tom Keel 2:31:34
The NERC supercomputer is the, what is, what, I was getting that, the Natural Environment Research Council.

Amulya Chevuturi 2:31:43
Just call it JASMIN.

Tom Keel 2:31:45
Yeah, yeah, but that's the JASMIN supercomputer that researchers use, and that keeps that seems to keep breaking as well, because just I think like running AI models is breaking a lot of compute and compute stuff.
But.

Yelesen, Sebastian 2:32:03

I think to be fair though, that's what's great about, like, that's what's good about Dask because it's kind of encouraged, like all of these good packages that are like encouraging people to think like, okay, don't worry that you don't have as much resource that you

Tom Keel 2:32:12

Yes.

YS

Yelesen, Sebastian 2:32:22

you might need, but actually, if you kind of use the resource you have in a smart way, you can achieve more than you thought, you know, and like, and yeah, but yeah.

Tom Keel 2:32:31

Exactly.

Yeah, Xarray and Dask is amazing for that. And then for tabular stuff, Polars is again, similar levels of.

Yeah, it's just so much quicker, basically, because it's closer to NumPy.

I'm going to jump in the other group, I think, because we've got 11 minutes left.

Going to see if this guy went up.

Amulya Chevuturi 2:33:04

Yeah. Anyone else, if they have questions, please do not hesitate to ask.

Hi, everyone. We will be bringing, we will be bringing the other group back in about 5 minutes, so you also have 5 minutes left.

Any questions, any queries?

They're coming back soon.

Matt Dalle Piagge 2:45:06

Oh my god, Teams did not crash. Amazing.

Like this is like the first time today it's done that. It's not crashed, I mean.

Amulya Chevuturi 2:45:12

What, what?

What are you guys doing that it did not crash?

Matt Dalle Piagge 2:45:19
I don't know; I never understand.
A lifelong learning experience for Microsoft Teams.

Amulya Chevuturi 2:45:25
They are updating it too fast to actually keep track of things.

Matt Dalle Piagge 2:45:29
Oh, well, that's half the problem, yeah.
Okay.

Tom Keel 2:45:34
Okay, so yeah, I think that's final 10 minutes. So we're just doing our wrap up. I can wrap this workshop up by saying thanks for everyone's participation. It's really lovely to see. And it was it was very tricky. I mean, we ran this last year and it had probably another 50% of content and it was really, really heavy. So I thought I'd cut it down, but apparently not. It's still quite heavy. Next year, it'll be a third of the size of that probably. But yeah, I'm hoping that everyone's taken something from this.
From this, I think the main thing is that as researchers, like we will have different ways of doing things and like it's basically all just comes down to setting out things nice and succinctly, nice and neatly, making sure you understand everything as you go along. So I'm hoping that
I've sort of demonstrated a short view of what it looks like when I my work, but yeah, maybe to other people that looks a little bit crazy. But yes, so is there any sort of final questions? Obviously we have 8 minutes before we finish, but just any final questions for the workshop or for any of the convenience?
And then, yeah, if not, you can always, if you get round to doing those independent tasks, you can always send us stuff or questions, but then we might have limited availability after this workshop ends. But yeah, please feel free to use this as a resource to.
And then, if not, I'll hand back over to Gianni or Amulya.

Amulya Chevuturi 2:47:37
Just to say that we will be around till 4:30, another 7 minutes. So feel free to ask us

questions afterwards also. But thank you for attending and we, or at least I will see you tomorrow. But off to you, Gianni then.

Tom Keel 2:47:41
Yeah, exactly, yeah.

Gianni Vesuviano 2:47:54
Yes, I'd like to thank all the participants for attending today. I'd like to thank the organisers of this workshop for putting this workshop on. And we'll see everybody tomorrow at the same time as we started today.
Have a nice evening or afternoon or tomorrow morning, depending on where you are in the world.

Tom Keel 2:48:18
Enjoy the football as well, if you're watching it.

Víctor Malagón Santos 2:48:18
Me too. Thank you so much.

Matt Dalle Piagge 2:48:21
Mm.

Víctor Malagón Santos 2:48:21
Yeah.
See you tomorrow.

Yelesen, Sebastian 2:48:23
Thanks for your time, everyone.

Gianni Vesuviano 2:48:24
Yeah.

Tom Keel 2:48:26
Thanks, bye.

Amulya Chevuturi 2:48:30

Tomorrow is funny.

Okay.

Matt Dalle Piagge 2:48:35

A good time zone to you all.

Amulya Chevuturi 2:48:37

Yeah.

◉ **Gianni Vesuviano** stopped transcription